REPORT RESUMES

ED 003 139 64

CURRICULUM GUIDE FOR A COURSE IN EDUCATIONAL MEDIA RESEARCH.

BY- STANLEY, JULIAN C. AND OTHERS

WISCONSIN UNIV., MADISON

REPORT NUMBER NDEA-VIIB-236 PUB DATE OCT 62

GRANT OEG-2-16-032

EDRS PRICE MF-$0.18 HC-$4.40 110P.

DESCRIPTORS- *CURRICULUM, *INSTRUCTIONAL MATERIALS, *MEDIA RESEARCH, *RESEARCH METHODOLOGY, STATISTICAL ANALYSIS, MADISON, WISCONSIN

A COURSE WAS DEVELOPED TO PREPARE CLASSROOM TEACHERS TO PERFORM ADEQUATE MEDIA RESEARCH. UPON COMPLETION OF THE COURSE, STUDENTS WOULD BE ABLE TO INITIATE, PERFORM, ANALYZE, AND REPORT RESEARCH THAT SHOULD PASS THE INSPECTION OF SPECIALISTS IN SUCH AREAS AS STATISTICS, PSYCHOLOGY, AND EDUCATIONAL MEDIA. THE COURSE WAS DESIGNED TO PRODUCE, ULTIMATELY, EXPERIMENTATION IN THE USE AND EVALUATION OF EDUCATIONAL MEDIA. TOPICS PLANNED FOR THE COURSE WERE--(1) GENERAL RESEARCH CONCEPTS, (2) RESEARCH DESIGN, (3) STATISTICAL ANALYSIS OF DATA, (4) DATA PROCESSING PROCEDURES, AND (5) RESEARCH REPORT WRITING. (HB)

ED003139

B-236

CURRICULUM GUIDE

for a course in

EDUCATIONAL MEDIA RESEARCH

RICHARD E. SCHUTZ
Arizona State University

ELLIS B. PAGE
University of Connecticut

JULIAN C. STANLEY
University of Wisconsin

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Title VII Project Number B-236
National Defense Education Act of 1958
Grant Number 2-16-032

The research reported herein was supported by a grant from the
U. S. Department of Health, Education, and Welfare
Office of Education
to the
University of Wisconsin; Julian C. Stanley, Chairman,
Department of Educational Psychology, Director

Curriculum Guide for a Course in Educational Media Research

Richard E. Schutz

Arizona State University

Ellis B. Page

University of Connecticut

Julian C. Stanley

University of Wisconsin

Prepared under USOE Title VII Contract B-236 with the University of Wisconsin; Julian C. Stanley, Chairman, Department of Educational Psychology, Director.

October 1962

Table of Contents

Page

I Introduction 1

II Behavioral Objectives of the Course 2

III Pertinent Pre-Course Experiences 3

IV Minimum Prerequisites 4

V Scope and Sequence of Course Content 6

VI Instructional Exercises 16

VII Suggested Reading 24

VIII Suggested Evaluative Procedures 26

IX Suggested Instructor Qualifications 38

Appendixes

A Experimental Design in Media Research

B Instructions for Preparation of Applications for Grants to Support Research Projects under the Provisions of Title VII of the National Defense Education Act of 1958

C Computer Program for Analysis of Variance

D Computer Program for Analysis of Covariance

I. Introduction

Courses in educational media and courses in educational research methods are both widely offered. Such courses are however, typically substantive: they talk about new media and about educational research methods. Few courses are designed to develop competence in media research. Those which are, usually have such a long list of prerequisites in audio-visual education, statistics, tests and measurements, etc., that they are beyond the reach of all but a small number of media specialists.

The course described here is designed to prepare competent classroom teachers to perform respectable media research. This does not mean, however, that the instruction attempts to cram five courses into one. Students completing the course will not be statisticians, psychologists, or even educational media specialists. But, if they achieve the objectives of the course, they will be able to initiate, perform, analyze, and report research that should pass the inspection of specialists in each of these areas. Although they will be novices as far as performance in research is concerned, they will be in position to perform. Moreover, their experiences in the course should predispose such performance and insure increased competency with further experience.

The ultimate objective of the course then is to produce experimentation in the use and evaluation of educational media.

In achieving this end we frankly hope to seduce the student to research efforts. We have tried hard to avoid the extremes of rape and nonproductive foreplay; the extremes which characterize many current research courses. The course does not assault the student with so much jargon and mathematics that he should find any further research efforts aversive. Yet neither is it a cozy but superficial "action research" course, which generates much enthusiasm but little else. Rather it gives the student direct experience in using powerful concepts and procedures basic to experimentation in educational media. The result we are expectantly awaiting is a fine crop of healthy legitimate media research studies.

II. Behavioral Objectives of the Course

When a person successfully completes the course he will be prepared to:

1. Define the expected outcomes of educational media usage in terms of observable behavior.
2. Locate previous research relevant to media usage, accurately interpret (explain) the results, and apply the findings in his use and evaluation of educational media.
3. Select the most valid experimental design for evaluating a media use that is practically feasible within the limits of his situation.
4. Perform the experimental operations consistent with the design selected.
5. Select the most appropriate statistical procedures for analyzing the experimental data.
6. Perform the necessary statistical analyses using the most appropriate computing vehicle (by hand, desk calculator, punch card processing, electronic data processing) depending upon the extent of the data or the complexity of the analyses.
7. Write up the results of his experiment in a standard form suitable for dissemination to various audiences.
8. Use the results of his experiment as a basis for further experiments.
9. Consult appropriate specialists for advice where needed at all stages of the research.

III. Pertinent Pre-course Experiences

Credentialed teachers can be expected to have had a number of common experiences which are directly relevant to the Media Research course. These include formal training in professional education and actual classroom teaching. Relevant courses commonly included in a teacher education program are:

1. Audio-visual instruction - Either in an entire course devoted to audio-visual media or in units of more general courses, students can be expected to become familiar with the function and use of media procedures and equipment.
2. Educational psychology - In one or more courses, students acquire concepts concerning learning and human development and are exposed to an operational approach appropriate to later conduct and evaluation of research.
3. Educational measurement and evaluation - Either in a single course or as part of methods courses, students acquire concepts of reliability, validity, prediction, standardization, etc., and become familiar with the commercially available tests.
4. Curriculum and/or methods courses - In one or more courses students acquire information concerning the sequencing of content and presentation of instruction.
5. In practice teaching and its related experiences students get supervised practice in the application of the concepts learned in the formal courses and get experience in using instructional media in the classroom. Further teaching in the field can be expected to broaden and deepen the practice teaching experiences.

IV. Minimum Prerequisites

Since this course is designed for able classroom teachers, the minimum prerequisites are those usually associated with teacher certification as listed above in Section III.

Although these experiences should assure prerequisite competency, successful research performance would seem to depend also on a degree of general ability commonly assessed by grade point average, success in graduate study, graduate level aptitude and achievement tests, and administrative ratings. Where the number of students must be restricted, it might be ill-advised to include a student not outstanding in these criteria who has also been teaching a considerable period with no observable creative contribution or professional growth.

The minimum qualifications required of students will vary depending upon the applicants available and the degree to which one really expects the students to achieve the course objectives. The higher the general caliber of the student the more likely will he produce generalizable and publishable research. In some situations it might be possible, and expedient, to be very selective in accepting applicants.

On the other hand, even if a student completing the course meets none of the objectives listed in Section II, his experience can still be worthwhile. Several objectives can be expected to be attained merely by "going through the motions" of the course. These include such items as:

1. An accepting attitude toward media experimentation.
2. General educational values of the knowledge of an unfamiliar discipline; i.e., an enlarged verbal repertoire.
3. Facilitation of communication between the teacher and the educational research specialist.
4. More willing and able participation of the teacher in larger experiments where he is responsible for a part of the replication.

For such purposes, insistence on unusual quality of teacher participants is not really necessary. In general, the competence implied by certification, together with some interest in the area of media research, would be evidence of meeting minimum prerequisites.

If enrollment is to be selective, however, several courses and experiences which cannot be expected to be included in the typical teacher education program might be recommended or desired background in selecting teachers who will achieve maximum understanding and application of the course content. Few applicants would likely have had many, if any, of these experiences. A person who has had them all would probably be in position to teach the course rather than take it. The following experiences then, are considered related. They should be noted but not expected.

A. Courses

1. Research methods
2. Elementary statistics or beyond
3. Differential psychology, social psychology, psychology of learning, psychology of perception, psychology of communication
4. Logic
5. Laboratory courses in various sciences
6. College mathematics

B. Professional experiences

1. Experience teaching a number of different subjects and grade levels
2. Experience in administration and/or supervision of instruction
3. Research assistantship
4. Experience in conducting personal research in connection with master's thesis, senior thesis, course projects, etc.
5. Coordinating or participating in school testing programs
6. Participating in committees for curriculum development or revision

V. Scope and Sequence of Course Content

The course will deal with the major topics outlined below:

A. General research concepts
   1. Sources for research ideas
   2. Framing experimental hypotheses
   3. Library search techniques
   4. Organization of a research prospectus

B. Research design
   1. Criteria for evaluating research designs
   2. Experimental designs
   3. Pre-experimental designs
   4. Quasi-experimental designs
   5. Correlational and ex post facto designs

C. Statistical analysis of data
   1. Taxonomy of experimental statistics
   2. Interpretation of statistical results

D. Data Processing procedures
   1. Functions of punch card equipment
   2. Card layout, coding, and flow charting
   3. Making use of computers

E. Writing the research report

Elaboration of these topics and a more detailed outline of the content of the course follows.

A. General Research Concepts

This initial section of the course will perform a dual purpose of building a repertoire of concepts basic to research activities and acquainting the students as rapidly as possible with current substantive developments in media research. Instruction will start at the top as far as the typical sequential conduct of a research study is concerned, considering first the potential procedures for generating worthwhile research ideas with brief attention to the current frontiers in media research. Next, procedures and established criteria for refining the problem and casting it into researchable form will be considered. Following this will come a treatment of the tactics of library search in locating background information relevant to a given problem. Finally, the section will culminate with a consideration of the purposes and characteristics of a research prospectus.

Thus, by the end of this first section of the course, the student should be reasonably up-to-date with what is currently going on in media research. Furthermore, he will be in position to move from a broad survey base to focus with increasing intensity on a narrower area of interest to him. Hopefully, he will have made some progress toward staking out a suitable research topic in this area.

A more detailed outline of this first section of the course follows.

1. Sources for research ideas
    a. Previous personal experience
    b. Previous course work
    c. Lists of research needs suggested by previous researchers
    d. Applying techniques or ideas developed in one context to a new situation to test their generalizability with respect to
        (1) Characteristics of subjects (age, geographical location, previous educational experience or achievement, etc.)
        (2) Content area (one school subject vs. another, in-class vs. out-of-school, etc.)
        (3) Research procedures (length or intensity of instruction, systematic modifications of materials, etc.)
    e. Finding a flaw in the logic or method of a published research study and replicating the study with the defect eliminated, controlled, or evaluated.

   - f. Opportune ideas, for which it is difficult to identify a specific antecedent source but which tend to occur only after one is thoroughly familiar with the published research in a given area.
2. Framing experimental hypotheses
   - a. Delimiting a problem
   - b. Operationally defining the terms used
   - c. Stating the problem in a testable form
3. Library search techniques - Including use of the following general reference sources
   - a. Library card catalog
   - b. *Education Index*
   - c. *Cumulative Book Index*
   - d. *Encyclopedia of Educational Research*
   - e. *Review of Educational Research*
   - f. *Psychological Abstracts*
   - g. *Dissertation Abstracts*
   - h. Research Abstracts in *AV Communications Review*
   - i. *Mental Measurements Yearbooks*
   - j. Inter-Library Loan procedures
4. Organization of research prospectus
   - a. Purpose of prospectus
   - b. Standard components of a prospectus

B. Research Design

This section of the course is intended to give the student the essential information he needs to select and utilize the most appropriate design for a given purpose. It is concerned then with the criteria for evaluating designs and the nature of the various designs which can be chosen and/or modified to suit a particular situation.

Appendix I presents material intended to provide the basic reference for this section of the course. This material is intended to serve as a convenient reference source, in open-book fashion. The student is not expected or advised to attempt to memorize or completely recall the contents of this material. The course experiences are designed to enable him to read it with understanding and use it as a functional tool in his future research efforts.

This section of the course, then, is concerned with the following topics:

1. Internal Validity
    a. History
    b. Maturation
    c. Testing
    d. Instrumentation
    e. Regression
    f. Selection
    g. Mortality
    h. Selection-Maturation Interaction
2. External Validity
    a. Testing and *X* Interaction
    b. Selection and *X* Interaction
    c. Reactive Arrangements
    d. Multiple - *X* Interference
3. Three True Experiments
    a. The Pretest-Posttest Control Group Design
    b. The Random Four-Group Design
    c. The Posttest-Only Control Group Design
4. More Complex True Experiments
    a. Interaction
5. Three Pre-Experimental Designs
    a. The One-Shot Case Study

   - b. The One-Group Pretest-Posttest Design
   - c. The Static Group Comparison Design
6. Quasi-Experimental Designs
   - a. The Time-Series Experiment Design
   - b. The Equivalent Time-Samples Design
   - c. The Equivalent Materials Design
   - d. The Nonequivalent Control Group Design
   - e. The Separate-Sample Pretest-Posttest Design
   - f. The Separate-Sample Pretest-Posttest Control Group Design
   - g. The Multiple Time-Series Design
   - h. The Recurrent Institutional Cycle Design
7. Correlational and Ex Post Facto Designs
   - a. Correlation and Causation
   - b. The Retrospective Pretest
   - c. Ex Post Facto Analyses

C. Statistical Analysis of Data

To analyze experimental data efficiently a researcher must know:

1. What statistical procedures to use that are appropriate for the data and that will provide a basis for inferences of interest.
2. How to use the results of these procedures in making inferences and in drawing conclusions based on these inferences.

In between these two steps is the computation. Statistical computations in this course will be treated as a necessary but mechanical aspect of research. The focus will be on 1 and 2 above: What statistics to use in connection with a design and how to interpret the results. It is expected that the student will "cook book" the computations if he performs them by hand or that he will get them performed by computer, using a standard program.

Table 1 presents a simplified taxonomy of experimental statistics intended as an aid in selecting appropriate statistical procedures. Each cell of the table directs the student to a source where he can find step-by-step computing directions.

Appendixes C and D present an example of a computer program description for the analysis of variance and the analysis of covariance, the two most widely useful statistical procedures.

Instruction in this section will include the following topics:

1. Use of the Simplified Taxonomy of Experimental Statistics
    a. Classification of data
        (1) Enumerated
        (2) Ranked
        (3) Metric
    b. Classification of samples
        (1) Matched
        (2) Random
    c. Using the reference as a guide in performing the computations
2. Interpretation of statistical results
    a. Levels of significance
    b. Errors in inference
    c. Statistical vs. practical significance
    d. Use of standard statistical tables
    e. Standard presentation of statistical results

Table 1

For each entry, a letter is given indicating one possible reference (see code below) and the first relevant page in the reference is indicated.

SIMPLIFIED TAXONOMY OF EXPERIMENTAL STATISTICS

Kind of Data	Kind of Samples				
		Two-sample		k-sample	
	One-sample	Matched	Random	Matched	Random
Enumeration	Binomial test (S, 36) (X^2 one-samp.) (S, 42)	McNemar Test for changes (S, 63)	Fisher exact-probability (s,96) X^2 test for two indep. samples (S, 104)	Cochran Q test (S, 161)	X^2 test for k indep. samples (S, 175)
Ranked	Kolmogorov-Smirnov one-sample runs test (S, 52)	Sign test (S, 68) Wilcoxon matched-pairs signed-ranks test (S, 75)	Median test (S, 111) Mann-Whitney U (S, 116) Kolmogorov-Smirnov two-samp. Wald-Wolfowitz runs test (S, 136) Moses test of extreme reaction (S, 145)	Friedman two-way analysis of variance (S, 166)	Extension of median test (S, 179) Kruskal-Wallis one-way analysis of variance (S, 184)
Metric		Matched t-test (E, 278) Two-way analysis of variance (E, 340) (E2, 158)	Random t-test (E, 246) One-way analysis of variance (E, 336) Analysis of covariance (E2, 281)	Two-way analysis of variance (E1, 340) (E2, 158)	One-way analysis of variance (E1, 314) Analysis of covariance (E2, 281)

Code: E1: A. L. Edwards. Statistical Methods for the Behavioral Sciences. New York: Rinehart, 1954.

E2: A. L. Edwards. Experimental Design in Psychological Research. New York: Holt, Rinehart, Winston, 1960.

S: S. Siegel. Nonparametric Statistics: for the Behavioral Sciences. New York: McGraw-Hill, 1956.

D. Data Processing procedures

Modern data processing equipment has taken nearly all of the "manual" labor ou of research. Computational difficulties no longer set much of a limit on the complexity of a research design or the number of subjects or replications involved. The researcher is free to design experiments on as "grand" a scale as he wishes without the unpleasant anticipation of spending a good portion of his life over a calculating machine analyzing the data. The treatment of design and analysis in Section B and C above is consistent with the new vistas made available by data processing machinery. The instruction there is designed to give the student the training he needs to locate efficiently the appropriate statistical procedures to analyze his data and interpret appropriately the results of his analyses. The in-between computations can either be turned over to a computer or handled mechanicall by the researcher working by hand to "simulate" the computer. Ordinarily, for an analysis of any size, it will be more efficient for a researcher to spend his time locating a computer, than to labor over a calculator.

To operate in this fashion the student must first be familiar with the capabilities of commonly used data processing machinery so that he will know what he can do with them. Secondly, he must be familiar with the common shop language used by data processing specialists so that he can establish communication with these people. This does not mean that he should be expected to acquire competence as a machine operator or computer programmer. Such specialists are available in adequat numbers and are usually willing to help out. However, data processing technicians cannot be expected to assume the responsibility for the researcher's work. Typically, they will lack the background and training necessary to help the experimenter choose appropriate statistical procedures or interpret analyses. This is the researchers' responsibility. However, if the researcher can describe what he wants done, in operational or procedural terms, the data processor can go ahead and do it. The instruction of this section of the course is therefore designed to enable the student to make efficient use of data processing men and machinery.

The following topics will be considered:

1. Functions of punch card equipment
    - a. Printing key punch
    - b. Card sorter
    - c. Reproducing punch
    - d. Card collator
    - e. Card interpreter
    - f. Alphabetic accounting machine

Most likely, because of their wide availability these data-processing machines will be products of International Business Machines Corporation. The functions of each machine will be described and the students will be given experience in performing standard operations on each machine.

2. Card layout, coding and flow charting. These are general data processing procedures which have wide application. Moreover they are procedures which the researcher must typically manage by himself since they must be taken care of at the planning stage of a research study prior to data collection. Excellent instructional materials listed in Section VII are available to assist in building competence in each of these areas, and exercises are outlined in Section VI to provide practical guided experience.

3. Making use of computers. This will involve first the development of a minimum repertoire of computer terminology. A second focus will be the use of standard statistical program libraries in general and individual program descriptions in particular. Finally, brief attention will be given to suggestions of potential computer installations and data processing services and specialists in each student's general locale to which he might reasonably look for assistance in connection with his future research efforts.

E. Writing the research report

This section of the course will be concerned with developing competence in using the standard style and form conventional in research writing. A distinction will be made between the form appropriate for academic theses, professional journals, and popular publication. The relative emphasis given to each of these forms of research reports will vary depending on the nature of the student groups and the general context of the course.

VI. Instructional Exercises

A. Students will be expected to engage in the conventional activities of "reading, attending lectures, contributing to discussions, taking examinations" as outlined in other sections of the Guide.

B. The "major project" for the course will be the individual preparation of a prospectus for a media research study using the Title VII proposal outline as a form guide (see appendix B). It is highly recommended that this assignment be broken down into smaller steps, with the sections being submitted sequentially, beginning as soon as possible in the course. Each section can then be critiqued by the instructor and/or other students. This practice avoids a good deal of misdirected time and effort on inappropriate problems, designs, procedures, etc.

C. Arrangements should be made to have the students visit local laboratories or field situations engaged in research in media or related fields to observe the nature of various research efforts at various stages of progress. The specific nature of the arrangements for observation and demonstration will vary with the resources present in a given locale.

D. Students will perform day-to-day directed exercises designed to give them experiences leading to the terminal behavior which constitutes the objectives of the instruction. These are described below.

Exercise on Defining Procedures and Objectives of Research

I. Define each of the following phenomena in terms of observable behavior indicators:

1. **The percepts generated by a sound motion picture**
2. **Teachers' appreciation of how the tape recorder can facilitate their accomplishments**
3. **Attitudes of authoritarian principals toward media usage**
4. **The values communicated by a televised cartoon series**
5. **Conscious interests of children as a basis for reducing prejudice**

Exercise on Library Search Techniques

List in detailed sequential order the reference sources you would consult to locate each of the following. Include the subject heading(s) you would look under where the reference includes an index.

1. Experimental research on methods of teaching spelling reported within the last twelve months.
2. A doctoral dissertation reporting an experiment on educational television written by someone at Indiana University about 1956.
3. A list of all books on self-instruction published to date.
4. A summary of work done in readability since 1950--to get an over-all view in less than 30 minutes.
5. A list of published tests which might have possibilities in evaluating achievement in arithmetic at the seventh grade level.
6. A comprehensive annotated bibliography on language laboratory equipment and utilization.

Exercise on Research Prospectus Wording

Each of the prospectus extracts below relates to an important point or set of points. However, it is faulty with respect to some technical aspect of procedure or presentation. Restate each extract, retaining the general validity of the point but presenting it more acceptably.

1. "In this experiment it is expected that there will be a greater probability that a difference will exist which will not be detected than that a difference will be detected due to a happenstance of chance rather than reality. Therefore, the largest possible sample and blocking should be used, and no special level of statistical significance will be required."

2. In a study investigating the effect of various techniques for teaching study skills, students are being randomly assigned to one of four treatment levels.

   "Some children who are taking extra subjects will be limited in the classes to which they can be assigned because of scheduling conflicts. Make special notations of those who could not be assigned to their first random assignment. These will be statistically treated later. Randomly assign them to the choices available to them."

3. "It is sometimes necessary, when trying to adopt methods of teaching which cost money, to prove the value of this new method over the present method and/or prove whereby this method would advance achievement enough to warrant its acceptance in the reading program. It is the latter which is promoting this investigation."

4. Hypothesis to be tested:

   "Neither of the two methods the Phonetic and the Decroly, will be proved to be more effective than the other one."

5. "The researcher believes that unquestionable results in behalf of the experimental group will be obtained. It should be noted a strong bias to some extent does exist and the researcher is aware of his bias. A .01 level of significance is needed to make the research of any true value."

6. "The experiment will be analyzed (Note: analysis is explained in adequate detail prior to this statement) so that it can be written up, published in some form, and results shared with the cooperating teachers and principals, school board and--particularly if results support the hypothesis--with teachers of other subjects for possible consideration."

Exercise on Punch Card Processing

Suppose you have the following situation:

Grade	Teacher	School	District	Number of Pupils
3	Mrs. Jones	Alpha	1	32
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
5	Mrs. Smith	Gamma	1	34
3 grades	50 teachers	5 schools	2 districts	550

There are no more than 10 teachers in any single school at a given grade or 40 pupils in any class.

You want to have 2 test scores punched on a card for each pupil. You plan to get:

1. A record of the 2 scores, listing each teacher's pupils in alphabetic order with the listings arranged by
   District
     Grade
       Teachers
2. A list of the scores on the first test arranged in ascending order by
   District
     Grade

Set up a code, and card layout and describe in sequence the data processing operations you would perform to obtain these listings. (You may use a flow chart to describe the operations if you wish.)

Exercise on Use of Punch Card Equipment

1. Keypunch the following data in the columns indicated

cols.	2-3	5-6	8-9	11-12	13-15	50-65	66-67
	28	26	27	17	423	HOYT	TA
	50	33	19	25	297	AARON	PR
	59	19	25	23	483	ABBOTT	PG
	52	41	29	30	692	ABDAI	J
	45	22	23	25	341	ADAIR	MJ
	62	28	29	28	151	ADKINS	OM
	43	20	15	21	573	ADLER	AE
	41	20	23	17	041	AKAMATSU	ME
	54	23	21	22	301	ALDEN	AS
	62	32	25	27	200	ALLEN	GM
	56	48	35	33	461	AYLOR	VC
	50	20	29	27	313	DUNCAN	BJ
	31	15	17	20	201	JOHNSON	JC
	63	28	26	24	137	SMITH	CA
	65	39	31	31	330	TAYLOR	DW
	36	27	24	20	513	TAYLOR	VL
	62	27	25	23	252	TEAGUE	JD

2. Reproduce the deck mechanically, retaining all the data but shifting the last name of each person from columns 50-65 to columns 40-55.
3. Gang punch a fixed identification code 621 in columns 76-78 of each card.
4. Interpret the duplicate deck.
5. Sort the original deck in alphabetic order.
6. Sort the duplicate deck in numerical order for the first two-column score.
7. List the contents of the two decks.

Exercise on Locating Computer Services

Consult the June issue of the journal, Computers and Automation titled "The Computer Directory and Buyer's Guide."

List the three sources that you would consult in each of the following categories. Rank the sources within each category in the order that you would contact them based on a combination of geographical location and services offered.

1. Commercial Computing Services
2. Commercial Consulting Services
3. College or University Computer Centers

VII. Suggested Reading

A. General research concepts

Any one or a combination of the following:

Good, C. V., and Scates, D. E. Methods of Research. New York: Appleton-Century-Crofts, 1954. Chapters 2 and 3.

Rummel, J. F. An Introduction to Research Procedures in Education. New York: Harper, 1958. Chapters 1-3.

Travers, R. M. W. An Introduction to Educational Research. New York: Macmillan, 1958. Chapters 1 and 4.

Van Dalen, D. B. Understanding Educational Research. New York: McGraw-Hill, 1962. Chapters 5-8.

Allen, E. M. Why are research grant applications disapproved? Science, 1960, 132, 1532-1534.

B. Research design (to be used in connection with Appendix A)

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963. Chapter 5, pages 171-246.

C. Statistical analysis of data (to be used in connection with table 1)

Edwards, A. L. Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1960.

Edwards, A. L. Statistical Methods for the Behavioral Sciences. New York: Rinehart, 1954.

McNemar, Q. Psychological Statistics. (3rd Ed.) New York: Wiley, 1962.

Siegel, S. Nonparametric Statistics. New York: McGraw Hill, 1956.

D. Data processing procedures

Borko, H. Computer Applications in the Behavioral Sciences. Englewood Cliffs, N. J.: Prentice Hall, 1962. Chapters 1-3.

IBM. The Design of IBM Cards. 25p. Form 22-5526-4.

IBM. General Information Manual: Introduction to IBM Data Processing Systems. 95p. Form F22-6517.

IBM. Modern Coding Methods. 26p. Form X21-3793.

IBM. Reference Manual: Flow Charting and Block Diagramming Techniques. 27p. Form 4-61:20M.

IBM. Reference Manual: Operator's Guide. 143p. Form 6-61:20M-W.

Postley, J. A. Computers and People. New York: McGraw-Hill, 1960.

Salmon, L. J. IBM Machine Operation and Wiring. Belmont, Calif.: Wadsworth, 1962.

E. Writing the research report

Any one or a combination of the following:

American Psychological Association, Council of Editors, Publication Manual of the American Psychological Association, 1957.

Campbell, W. G. Form and Style in Thesis Writing. Boston: Houghton Miffin, 1954.

Trelease, S. F. How to Write Scientific and Technical Papers. Baltimore: Williams and Wilkins, 1958.

Turabian, K. S. Manual for Writers of Term Papers, Theses, and Dissertations. Chicago: University of Chicago Press, 1960.

VIII. Suggested Evaluative Procedures

At the end of the course a student should be able to make the responses specified in the course objectives without special cues or prompts. This implies: (a) that the responses are the student's own and not those of some classmate and (b) that the responses are generalized or transferred from the illustrations and conditions of the workshop to other situations resembling those in which he may be later working. Meeting these requirements involves evaluative procedures.

On the following pages are illustrative exercises for evaluating performance in each major area of the course. The exercises are divided into three "one-hour" examinations. The exercises may, however, be "cut and pasted" to meet different time limits or objectives.

Test Over Chapter Entitled

"Experimental and Quasi-Experimental Designs for Research on Teaching"*

Instructions: For each question, pick the best answer from among those given. This will at times require the study of the accompanying panel of experimental designs. You should note that the design numbers suggested in the answers are not those design numbers used in Campbell and Stanley, but those on the panel. Mark the letter of your answer in the appropriate space on the answer sheet.

There is no negative weight for incorrect choices. When you do not know an answer, you should make the best guess.

1. Which one of the following designs is a "true" experimental design, according to Campbell & Stanley?
A. 1 B. 2 C. 3 D. 4 E. 5

(Note: Decide from the diagrams which of those numbered 1, 2, 3, 4, or 5 is the best answer. Then mark the letter associated with that number on the answer sheet as your answer to the first question.)

2. Which of the following is a true experimental design?
A. 6 B. 7 C. 8 D. 9 E. 10

3. Which of the following is a true experimental design?
A. 11 B. 12 C. 13 D. 14 E. 15

4. Which of the following is the "Solomon Four-Group Design"?
A. 1 B. 6 C. 9 D. 12 E. 14

5. Which is termed a "Counterbalanced Design"?
A. 1 B. 6 C. 9 D. 12 E. 15

6. Which of the following designs is most subject to invalidity due to "history"?
A. 8 B. 10 C. 11 D. 13 E. 15

7. Which represents the "institutional cycle design"?
A. 2 B. 8 C. 10 D. 14 E. 15

---

*Which is Chapter 5, pages 171-246, in N. L. Gage (Editor), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

8. The advantage of design (1) over (4) is the comparative absence in (1) of the danger of

   A. instrumentation.
   B. interaction of testing and X.
   C. interaction of selection and X.
   D. reactive arrangements.
   E. multiple-X interference.

9. A design not pictured in the panel is

   A. regression discontinuity.
   B. separate-sample pretest-posttest control group.
   C. multiple time series.
   D. nonequivalent control group design.
   E. one-shot case study.

10. The advantage of (11) on the panel compared with (5) is the avoidance of much of the danger of

    A. history.
    B. interaction of testing with X.
    C. interaction of maturation with X.
    D. selection biases.
    E. regression.

11. Which one of the designs in the panel is an example of what McCall called a "rotation experiment"?

    A. 6 B. 8 C. 10 D. 12 E. 14

12. Where one cannot control the treatment but can control the observation, employing random assignment to the measurements, then one may wish to use which design of the panel?

    A. 2 B. 3 C. 6 D. 13 E. 15

13. The particular advantage of design (10) in the panel over design (8) is that some attempt is made to control

    A. history.
    B. selection.
    C. mortality.
    D. multiple-X interference.
    E. interaction of selection and maturation.

14. The "equivalent time-samples" design is pictured in

    A. 2 B. 8 C. 10 D. 12 E. 14

15. Because of its particular nature, the design (4) in the panel is

    A. well worth using much of the time.
    B. seldom worth using because generalizations are difficult to make.
    C. never worth using.
    D. a true experiment.

16. **Which one of the following is not usually considered a threat to the internal validity of an experiment?**

    A. Maturation
    B. Testing
    C. Reactive arrangements
    D. Regression
    E. Mortality

17. **A design widely accepted in the physical sciences but of relatively poorer status in the social sciences is**

    A. 3 B. 7 C. 8 D. 11 E. 15

18. **If you had only one intact classroom with which to experiment, a design which might be useful would be (best choice)**

    A. 2 B. 8 C. 10 D. 12 E. 13

19. **Annual achievement tests, illness records, daily assignments, and other routine characteristics and activities of the school are considered, for experimental purposes,**

    A. reactive B. nonreactive

20. **Of the following, the design with least external validity would typically be**

    A. 1 B. 2 C. 5 D. 11 E. 12

21. **Regression-discontinuity analysis, according to Campbell & Stanley,**

    A. is almost a true experiment.
    B. is preferable to random division of marginal scores, into award and non-award.
    C. is subject to invalidity due to history.
    D. is subject to invalidity due to selection-X interaction.
    E. should almost never be employed.

22. **Number (3) in the panel is not a "true" experiment because**

    A. it does not provide for pretest control.
    B. it does not provide for random assignment to treatment groups.
    C. there is no control of instrumentation.
    D. there is no control of maturation.
    E. A-D are wrong; it is a true experiment.

23. **Ex post facto designs are judged by the authors to be**

    A. useful when other analysis is difficult.
    B. true experiments under certain narrow conditions.
    C. subject uniquely to instrumentation errors.
    D. unsatisfactory at best.

24. **Design (10) is judged typically weakest in**

    A. history B. maturation C. testing D. selection
    E. interaction of testing and X

25. Suppose you are studying the effect of "emotionality" in a speech. For purposes of generalization of X, which is preferable? Use

A. a tape recorder with the speech set.
B. one speaker, delivering same speech in same way.
C. different speakers delivering speech in practiced way.
D. different speakers interpreting speech in a variety of ways.

26. Which of the following designs best controls for biasing in selection?

A. 1 B. 2 C. 8 D. 10 E. 11

27. In classroom experiments where achievement in a subject is the criterion, which solution best answers the problem of biased assignment and unique intragroup histories?

A. Moving the randomization to classroom as unit.
B. Setting up randomization procedures to assign students to classes at the beginning of fall semester.
C. Blocking on the basis of individual student ability.
D. Applying analysis of covariance to individual student ability.
E. Use of the within-class variance as error term.

28. For educational experiments where randomization is possible, what is the preferred policy regarding pre-tests which are related closely with the post-tests?

A. Almost always preferred
B. Usually preferred
C. Usually not preferred

29. Where random assignment is not possible but groups are tested in situations where "nature" has run the experiments, the experimenter has the duty of

A. denying any causal relationship in his data.
B. examining any rival explanations of the data.
C. stating his assumption that any other influences are randomly distributed among "treatments."
D. finding other educational problems to investigate.

30. Suppose we have 30 subjects and we roll a die repeatedly to assign them to three groups. All study particular assignments and take a test at the end of the study period. One group studies under a condition of music, one under a condition of random industrial noise, and one under silence. Which of the designs are we using?

A. 1 B. 3 C. 4 D. 5 E. 11

31. Suppose in the experiment above we could not separate these randomly but had to use three classes which already existed. Now which design would it be?

A. 1 B. 3 C. 4 D. 5 E. 11

32. In the experiment in (30) the individual subject is which of the following?

A. Nested
B. Crossed

33. Suppose you need to run an experiment in which one of the treatments (a field trip) requires the permission of parents. What is the best procedure?

A. Have alternate students standing by to replace any who are unable to secure permission of parents in that treatment group.

B. Let the students go but assess the loss by dropping those students who seemed like the no-shows in the relevant variables.

C. Secure permission from all possible parents before the experiment, and before assignment of students to treatments.

D. Assume that drop-out is random and disregard dangers in the analysis.

34. Suppose you have two intact classes available, both taught by the same teacher, and wish to test the efficacy of a particular visual aid in instruction. For each unit through the semester the aid is assigned randomly to one class or the other. At the end of a semester, the units taught by visual aid are significantly superior to those taught without. You can generalize the results to presence of the aid in general under which of the following assumptions?

A. No interaction exists between treatments and any of the specific variables such as teacher, or class, etc.
B. This teacher is an average teacher and the class is an average class.
C. No interaction exists between maturation and history.
D. The classes, although intact, are essentially equivalent.

35. Suppose you wish to "test" a favorite theory, and construct an experiment to do so. From your findings you are able to reject the null hypothesis at the .01 level. What may you conclude? (best choice)

A. You have validated the theory.
B. Your findings support the theory.
C. The theory has not been disconfirmed.
D. The theory has not been supported.

KEY

1 - E
2 - D
3 - A
4 - C
5 - D
6 - D
7 - E
8 - B
9 - A
10 - B
11 - D
12 - C
13 - A
14 - A
15 - A
16 - C
17 - C
18 - A
19 - B
20 - B
21 - D
22 - B
23 - D
24 - E
25 - D
26 - E
27 - A
28 - C
29 - B
30 - E
31 - B
32 - A
33 - C
34 - A
35 - C

Panel of Designs to Accompany Test

The student should already understand the following diagrams and their notation. X is some experimental treatment. R is the random assignment to groups. O is some observation of behavior. A broken line indicates the use of intact groups without random assignment. Numbers here assigned to the designs are for use in answering problems in the quiz, and have no other meaning. They are not the design numbers from the text.

1.
$$\begin{array}{llll} R & O & (X) & \\ R & & X & O \\ \hdashline R & O & & \\ R & & & O \end{array}$$

2. $X_1O \; X_0O \; X_1O \; X_0O$ etc.

3.
$$\begin{array}{ll} X & O \\ \hdashline & O \end{array}$$

4.
$$\begin{array}{lll} O & X & O \\ \hdashline O & & O \end{array}$$

5.
$$\begin{array}{llll} R & O & X & O \\ R & O & & O \end{array}$$

6.
$$\begin{array}{llll} R & O & (X) & \\ R & & X & O \\ & & (X) & \\ & & (X) & \end{array}$$

7. X O

8. $O_1 \; O_2 \; O_3 \; X \; O_4 \; O_5 \; O_6$

9.
$$\begin{array}{llll} R & O & X & O \\ R & O & & O \\ R & & X & O \\ R & & & O \end{array}$$

10.
$$\begin{array}{lllllll} O & O & O & X & O & O & O \\ \hdashline O & O & O & & O & O & O \end{array}$$

11.
$$\begin{array}{lll} R & X & O \\ R & & O \end{array}$$

12.
$$\begin{array}{llll} X_1O & X_2O & X_3O & X_4O \\ \hdashline X_2O & X_4O & X_1O & X_3O \\ \hdashline X_3O & X_1O & X_4O & X_2O \\ \hdashline X_4O & X_3O & X_2O & X_1O \end{array}$$

13. O X O

14. $M_aX_1O \; M_bX_0O \; M_cX_1O \; M_dX_0O$ etc.

15.
$$\begin{array}{lllll} \text{Class A} & X & O_1 & & \\ \hdashline \text{Class B} & & O_2 & X & O_3 \end{array}$$

Statistical Usage Examination

I. Below are brief descriptions of possible research problems. For each situation list the most appropriate statistical technique(s) for analyzing the data involved.

1. A questionnaire survey indicates that 83% of audio-visual directors and 62% of film library supervisors read the *AV Communications Review* regularly. 143 questionnaire reports are available for each group. Do these data indicate a difference in readership?
2. An investigator wishes to determine if there is any difference in the spatial relations ability of 640 boys and girls who completed a 3500 frame self-instructional program on "How to Take a Test."
3. A researcher wishes to determine the relationship between scores on the NEW Test of Creativity (range = 0 - 64) and the OLD Intelligence Test (IQ range 94 - 116).
4. 45 of 105 secondary teachers in a unified school district rated a film they had just viewed as "warm" on a forced choice item "warm - aloof" while 30 of 73 elementary teachers rated the film "warm". Is there a difference in the two groups?
5. A researcher shows a film in black-and-white and in color. He wishes to determine if the addition of color has any effect on achievement as measured by a standardized test.

II. Below is an abbreviated outline of a study to compare three feedback techniques in television teaching (*AV Communications Review*, 1962, 10, A - 100-101.)

Subjects: 80 students were selected at random from the total full-time undergraduate enrollment at the American University and randomly assigned to one of four groups. In each group, 12 subjects completed the experiment and were used in the analysis.

Procedure: Televised instruction was presented under four conditions: (1) To Group I without feedback; (2) To Group II with vicarious feedback; (3) To Group III with feedback through a classroom microphone; (4) To Group IV with electrical signal feedback. Each of the four groups was exposed to five 50-minute televised lectures dealing with the theory and application of Group Dynamics.

Test Measures: Immediately after the fifth lecture, the subjects were given a test on the content of the lectures and they also filled out a questionnaire on their attitudes toward TV instruction.

Prepare the first part of the necessary analysis of variance table(s) showing the sources of variation and degrees of freedom for each analysis of these data you would make.

III. A researcher is interested in knowing whether self-instructional programs are more effective in teaching algebra than straight lecture and workbook. He also wants to know whether it makes a difference which school the student attends. He sets up an experiment involving three variations of teaching machine and traditional method in four different schools. His analysis of variance of the results of the study is presented below:

Source of Variation	df	MS	F
Between Methods	2	24,365.22	5.34*
Between Schools	3	3,158.92	.69
M x S	6	13,840.63	3.03*
Error	12	4,563.22	

* $p < .05$

Interpret these findings

Data Processing Examination

(Students should be provided with a card layout sheet. One possible format for such a sheet is shown on page 37. They will also need access to the analysis of variance program description included in Appendix C.)

I. Prepare a card layout for coding and key punching the data included on the questionnaire excerpt on the following page.

II. Prepare a flow chart showing the operations necessary to:

1. Count the responses to each item
2. List and punch the summary tabulations
3. Prepare a duplicate deck of the original punched responses, printed so that the information can be read directly from the cards.

III. Suppose you wished to perform a 3 x 4 x 4 complete factorial analysis of variance with 20 replicates. Show on your card layout sheets the necessary card preparation for running these data on the Analysis of Variance program.

Questionnaire Extract

1. How old were you when you applied for the first N.D.E.A. loan you received?

Below 17____ 22____
17____ 23____
18____ 24____
19____ 25____
20____ 26____
21____ 27 or over ____

2. What is your sex?

Male____ Female____

3. What was your marital status when you applied for your first loan?

Married____ Widow(er)____
Single____ Divorced____

4. For how many dependents were you responsible at the time of your first application for a N.D.E.A. loan?

0____ 4____
1____ 5____
2____ 6____
3____ Over 6____

5. If you had children at the time you applied for your first loan, what was the age of your oldest child?

0 - 1 years____ 5 - 6 years____
1 - 2 years____ 6 - 7 years____
2 - 3 years____ 7 - 8 years____
3 - 4 years____ 8 - 9 years____
4 - 5 years____ 9 -10 years____

6. What age were you when you decided to become a teacher?

Under 10____ 21 - 25____
11 - 15____ 26 - 30____
16 - 20____ Over 30____

7. Did your parents or guardians approve of your selection of teaching as an occupation?

Father Mother
Yes____No____ Yes____No__

8. How many brothers and sisters did you have the year you applied for your first loan?

0____ 5____
1____ 6____
2____ 7____
3____ 8____
4____ over 8____

9. When you applied for your first loan, in which grade were you?

High School Senior____ College Junior____
College Freshman____ College Senior____
College Sophomore____

10. When you applied for your first loan, had you already been enrolled in a professional education class?

Yes____ No____

11. How many years of military service had you completed when you applied for your first loan?

0____ 3____ 6____
1____ 4____ 7____
2____ 5____ Over 7____

Card Layout Sheet

Job ______________________ Date __________ Card No. __________

Col.

1 ______________________
2 ______________________
3 ______________________
4 ______________________
5 ______________________
6 ______________________
7 ______________________
8 ______________________
9 ______________________
10 ______________________
11 ______________________
12 ______________________
13 ______________________
14 ______________________
15 ______________________
16 ______________________
17 ______________________
18 ______________________
19 ______________________
20 ______________________
21 ______________________
22 ______________________
23 ______________________
24 ______________________
25 ______________________
26 ______________________
27 ______________________
28 ______________________
29 ______________________
30 ______________________
31 ______________________
32 ______________________
33 ______________________
34 ______________________
35 ______________________
36 ______________________
37 ______________________
38 ______________________
39 ______________________
40 ______________________
41 ______________________
42 ______________________
43 ______________________
44 ______________________
45 ______________________
46 ______________________
47 ______________________
48 ______________________
49 ______________________
50 ______________________
51 ______________________
52 ______________________
53 ______________________
54 ______________________
55 ______________________
56 ______________________
57 ______________________
58 ______________________
59 ______________________
60 ______________________
61 ______________________
62 ______________________
63 ______________________
64 ______________________
65 ______________________
66 ______________________
67 ______________________
68 ______________________
69 ______________________
70 ______________________
71 ______________________
72 ______________________
73 ______________________
74 ______________________
75 ______________________
76 ______________________
77 ______________________
78 ______________________
79 ______________________
80 ______________________

IX. Suggested Instructor Qualifications

The course may be fielded by a single instructor or by an instructional team. Whether an individual or an instructional team is used will depend on such considerations as:

1. Whether necessary competencies (see below) are obtainable in an individual or only through a number of individuals.
2. Whether the course is highly concentrated. If students are in class five hours a day, for example, more than one instructor would seem desirable.
3. The number of students or participants. Beyond a certain number, more than one instructor would seem indicated.
4. The heterogeneity of the students. If students differ widely in achievement or experience, more than one section will likely be required.
5. The objectives of the course. If purposes are broadly conceived as general education (see Section IV above), then a higher participant-professor ratio may be tolerated than where purposes are conceived as producing creative individual work during the course.
6. The amount of personalized application expected. With considerations related to 5 above, if the course is expected to result in individual designs in media research, more professional aid is clearly required than if it is not.

In choosing an instructor (or instructors) for a course of this type, one must always settle for the nearest approximation to the ideal that is practically obtainable. But the following criteria would appear useful in considering such an appointment:

1. An achieved doctoral degree in the psycho-educational area.
2. Interest and competence in educational research (preferably demonstrated by respected published research).

3. Particular competence in either or both of the following fields (depending on whether one or more instructors are scheduled for the workshop):
   a. Experimental design, with graduate training at least through courses in experimental design, and competence in such areas as analysis of variance and covariance, regression techniques, factor analysis and non-parametric tests, etc. Should have interest in and enthusiasm for design questions.
   b. Media research, with experience and ability (preferably demonstrated by publication) in educational media research and knowledge of the traditional and the recent investigations of such media. Preferably this person should have professional contacts with media researchers throughout the United States in order to know the major lines of current investigation. Practical experience in media operations probably is of minor importance unless such experience has been research-oriented. Ideally, this person will be aware of broad psychological questions implicit in media investigation.

APPENDIX A

Experimental Design in Educational Media Research

[Based on Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research on Teaching. Chapter 5 (pages 171-246) in Gage, N. L. (Ed.), Handbook of Research on Teaching (A project of the American Educational Research Association - NEA). Chicago: Rand-McNally, 1963.]

Experimental Design in Media Research

Table of Contents

Page

Internal Validity 1
    History 2
    Maturation 2
    Testing 2
    Instrumentation 3
    Regression 3
    Selection 6
    Mortality 6
    Selection-Maturation Interaction 6
External Validity 7
    Testing and X Interaction 7
    Selection and X Interaction 8
    Reactive Arrangements 8
    Multiple - X Interference 10
Three True Experiments 11
    1. The Pretest-Posttest Control Group Design 11
    2. The Random Four-Group Design 14
    3. The Posttest-Only Control Group Design 14
More Complex True Experiments 16
    Interaction 16
Three Pre-Experimental Designs 19
    4. The One-Shot Case Study 19
    5. The One-Group Pretest-Posttest Design 20
    6. The Static Group Comparison Design 20

Page

Quasi-Experimental Designs 21

7. The Time-Series Design 21

8. The Equivalent Time-Samples 25

9. The Equivalent Materials Design 26

10. The Nonequivalent Control Group Design 27

11. The Separate-Sample Pretest-Posttest Design 29

12. The Separate-Sample Pretest-Posttest Control Group Design 31

13. The Multiple Time-Series Design 32

14. The Recurrent Institutional Cycle Design 33

Correlational and Ex Post Facto Designs 36

Correlation and Causation 36

The Retrospective Pretest 39

Ex Post Facto Analyses 40

Summary 42

Tables

Page

1. Sources of Invalidity for Designs 1 through 6. 15

2. Sources of Invalidity for Quasi-Experimental Designs 7 through 14 . . . . 23 & 24

Figures

Page

1. Regression in the prediction of posttest scores from pretest, and vice versa 4

2. Some possible outcomes of a complex experiment 17

3. Some possible outcomes from the introduction of an experimental variable at point X into a time-series of measurements 22

This presentation examines the validity of 14 experimental designs against 12 common threats to valid experimentation. In presenting the experimental designs the following code is used:

X = the exposure of a group to an experimental event or
treatment of interest (E.g., some form of media.)

O = some process of observation or measurement
(E.g., a checklist, rating scale or test.)

The *X*s and *O*s in a given row are applied to the same persons, with temporal order indicated by left to right sequence. Thus, if a group were exposed to a film, and then tested, the procedures would be coded as:

X *O*

*X*s and *O*s vertical to one another are simultaneous. Thus, if we were to add another post-test group to the group above, the situation would be coded as:

X *O*
O

R = random assignment to treatment groups.

Such assignment is made prior to the initial treatment or experimental observation. *Randomization is the all-purpose procedure for achieving approximate pre-treatment equality of groups* by ruling out systematic bias in assigning experimental units (e.g., students) to treatment groups.

Internal Validity

Internal validity is the basic minimum without which any experiment is uninterpretable. It asks the question: Did in fact the experimental treatments make a difference in this specific experimental instance? Eight "threats" to internal validity will be presented here. Each of these is an extraneous influence which, if not controlled in the experimental design, may produce results mistakenly attributed to the effect of the experimental treatment. The eight threats are:

History: other events occurring between a first and second measurement in addition to the experimental treatment. Consider the sequence O_1 X O_2. Between O_1 and O_2 many other change-producing events may occur in addition to the experimenter's X. If the pretest (O_1) and the posttest (O_2) are on different days, for example, some naturally-occurring event in between may have caused the difference. To become a *plausible* rival hypothesis, such an event should have occurred to most of the students in the group under study, say in some other class period or via a widely disseminated news story. History becomes a more plausible rival explanation of change the longer the O_1 - O_2 time lapse, and might be regarded as a trivial problem in an experiment completed within a one or two hour period, although even here, such things as laughter, distracting events, etc., can have strong effects.

Maturation: biological or psychological processes which systematically vary with the passage of time per se, independently of the experimental treatment or other events. Thus between O_1 and O_2 students may grow older, hungrier, more tired, more bored, etc., and the obtained difference may reflect this rather than X. In remedial education of exceptionally disadvantaged persons, for example, a process of "spontaneous improvement", analogous to the natural healing of wounds, may be mistaken for the effect of a remedial X. (Needless to say, such an improvement is not regarded as "spontaneous" in any causal sense, but rather represents the cumulative effects of experiences which occur even if no X is introduced.)

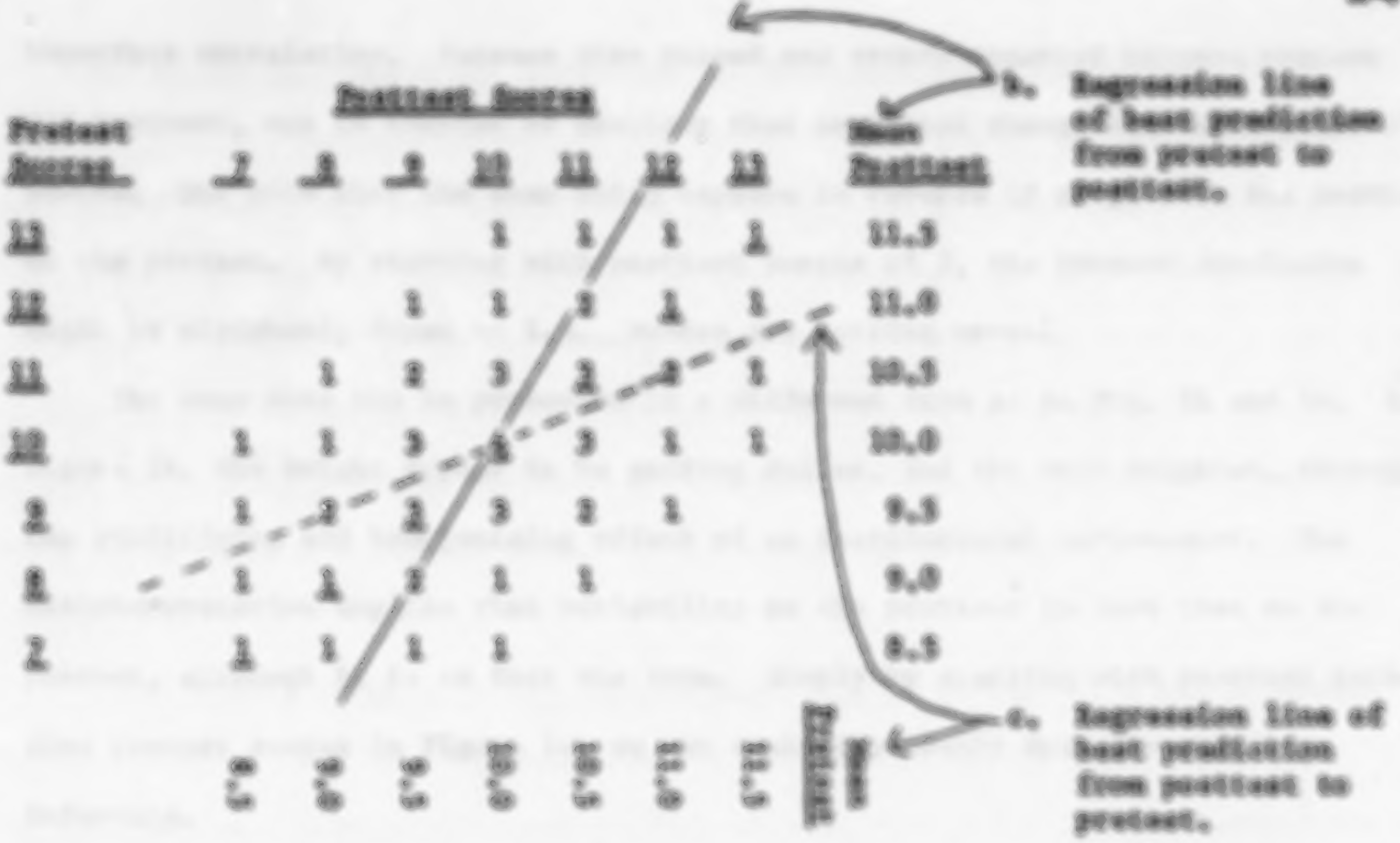
Testing: the effects of a pre-test on subsequent behavior. Students taking achievement and intelligence tests a second time usually do better than those taking the test for the first time. These effects occur without any special instruction about the test. The pre-test experience itself accounts for the improved test score. A similar effect is often noted for personality tests. On a second test an individual tends to perform as if he were better "adjusted."

Instrumentation: changes in the measuring instrument itself, or in the observers or raters, which might account for an $O_1 - O_2$ difference. These changes are analogous to the stretching or fatiguing of a spring scale, condensation in a cloud chamber, etc. Where human observers are used to obtain O_1 and O_2, changes in their behavior due to learning, fatigue, etc., will produce $O_1 - O_2$ differences. If essays are being graded, for example, grading standards may shift between O_1 and O_2 (suggesting the control technique of shuffling the O_1 and O_2 essays together and having them graded without knowledge of which came first). If classroom participation is being observed, then the observers may become more skillful, or more blasé, on the second occasion. If parents are being interviewed, the interviewer's increased familiarity with the interview schedule and with the particular parents may produce shifts. The use of different observers or raters for O_1 than for O_2 could also cause a difference.

Regression: the universal tendency of extreme scores to be closer to the mean on a second test. Suppose, for example, students are picked for special experimental treatment because they do particularly poorly on an achievement test (which becomes for them the O_1). Their scores on the same or a parallel test given later at O_2 will almost surely average higher than their O_1 scores. This dependable result is not due to an effect of X or a practice effect. It is rather an artifact of the less-than-perfect correlation between O_1 and O_2.

Figure 1 shows what happens when a pretest and a posttest for a whole population correlate .50, with no change in mean or variability.

In this example, no true change has taken place, i.e., the pre- and post-test mean scores are the same. But looking at the posttest performance of individuals scoring very low on the pretest, e.g., scores of 7, one finds their posttest scores scattered, but in general better. They have on the average "regressed" one half way back to the group mean (i.e., correlation coefficient is .50), resulting in an average of 8.5. But rather than representing progress, this is totally a result of

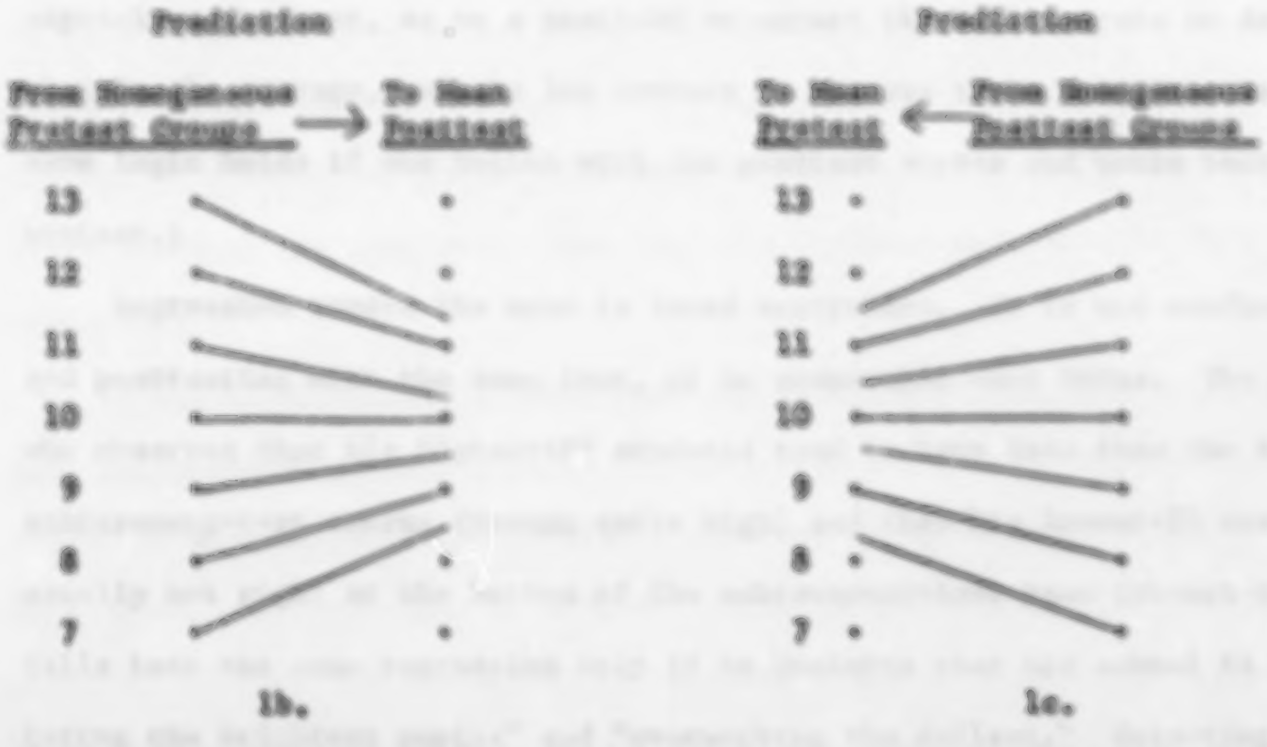


Pretest Scores	7	8	9	10	11	12	13	Mean Posttest
13				1	1	1	1	11.5
12			1	1	2	1	1	11.0
11		1	2	3	3	2	1	10.5
10	1	1	3	4	3	1	1	10.0
9	1	2	3	3	2	1		9.5
8	1	1	2	1	1			9.0
7	1	1	1	1				8.5
Mean Pretest	8.5	9.0	9.5	10.0	10.5	11.0	11.5	

b. Regression line of best prediction from pretest to posttest.

c. Regression line of best prediction from posttest to pretest.

1a. Frequency scatter of posttest scores for each class of pretest scores, and vice versa.


Prediction
From Homogeneous Pretest Groups
To Mean Posttest
13
12
11
10
9
8
7
Prediction
To Mean Pretest
From Homogeneous Posttest Groups
13
12
11
10
9
8
7

1b.

1c.

Figure 1. Regression in the prediction of posttest scores from pretest, and vice versa.

imperfect correlation. Because time passed and events occurred between pretest and posttest, one is tempted to conclude that some real change has taken place in scores. But note that the same thing happens in reverse if we go from the posttest to the pretest. By starting with posttest scores of 7, the reverse conclusion might be mistakenly drawn -- i.e., scores are getting worse!

The same data can be presented in a different form as in Fig. 1b and 1c. In Figure 1b, the bright appear to be getting duller, and the dull brighter, through the stultifying and homogenizing effect of an institutional environment. The misinterpretation implies that variability on the posttest is less than on the pretest, although it is in fact the same. Simply by starting with posttest rather than pretest scores in Figure 1c, we can again mistakenly draw the opposite inference.

We may look at regression toward the mean in another, related way. In a sense the typical extremely high scorer has had unusually good "luck" (large positive error) and the extremely low scorer bad luck (large negative error). Luck is capricious, however, so on a posttest we expect the high scorers to decline somewhat on the average, and the low scorers to improve their relative standing. (The same logic holds if one begins with the posttest scores and works back to the pretest.)

Regression toward the mean is found everywhere. It is not confined to pre- and posttesting with the same test, or to comparable test forms. The principal who observes that his highest-IQ students tend to have less than the highest achievement-test scores (though quite high) and that his lowest-IQ students are usually not right at the bottom of the achievement-test heap (though quite low) falls into the same regression trap if he declares that his school is "understimulating the brightest pupils" and "overworking the dullest." Selecting those students who scored highest and lowest on the achievement test and looking at their IQs would force him, by the same illogic, to conclude just the opposite!

Selection: *Pre-treatment inequality of experimental groups*. When individuals are assigned to treatment groups on the basis of availability or convenience or even matched on background characteristics other than *O*, there is no formal means of certifying that the groups would have been equivalent had it not been for *X*. If after the exposure to *X* the two groups differ, this difference could well have come about through the differential assignment of persons to the groups: the groups might have differed anyway, without the occurrence of *X*. Let us repeat: Randomization is the only all-purpose procedure for achieving pre-treatment equality of groups.

Experimental mortality: *the differential loss of persons from the treatment groups*. Even if two non-randomly assigned groups were once alike, they might differ at the time of the experiment not because of a change in the *members* themselves, but in the *membership*. In educational research, for example, this problem frequently arises in studies of the effects of a college education which compare freshmen (who have not had the *X*) with seniors (who have). When freshmen women are shown to be more beautiful than senior women, we need not recoil at the implication that our harsh course of training is debeautifying. Instead, we may point to the hazards in the way of a beautiful girl's finishing college before getting married. (Of course, if we consider the *same* girls when they are freshmen and seniors, this problem disappears.)

Selection-Maturation Interaction: *the effect obtained when treatment groups have different maturation rates*. This differential rate of maturation obscures the effect of any treatment *X*. Suppose a group of delinquents is tested, experimentally treated, and retested. A "control" group of non-delinquents is tested but not treated. If only the delinquent group shows a gain, this might well be spontaneous improvement occurring only in an extreme group. Such a gain would have occurred in the delinquents even without *X*, and would represent selection-maturation interaction, if it could be demonstrated that one group is more "mature" than the other.

External Validity

External validity asks the question of *generalizability*: to what populations, settings, treatments, and measurements can an effect be generalized? While this question, like all general questions of inductive inference, can never be completely answered, it can best be solved by designs which are valid both internally and externally. Factors of internal validity produce direct changes in *O* scores which might be mistaken for the results of *X*. Threats to external validity on the other hand involve a *differential* interaction of *X* with some other variable. They thus threaten the ability to generalize to other groups. These threats to external validity are the following:

Testing and X Interaction: In which a pretest increases or decreases individual sensitivity or responsiveness to the treatment and thus makes the experimental results for pretested individuals different from what they would have been without the pretest. A reactive effect of this sort can be expected whenever the test itself is a stimulus to change, rather than a passive record of behavior. Thus in an experiment on education for weight control, the initial weigh-in might itself lead to diet control and weight reduction, without any educational treatment. Similarly, placing observers in a classroom to observe a student teacher's pre-training skills may itself change the teacher's later classroom conduct. Placing a microphone on the desk may change the group interaction pattern, etc. In general, *the more novel and motivating the test device, the more reactive it will probably be*.

Paper-pencil testing, however, is a regular classroom event. Especially if an experiment uses regular or conventional classroom examinations as *O*s, the undesirable *testing* and *X* interaction is certainly reduced. Where highly unusual tests are used, or where the testing procedure involves surprise, stress, etc., it is better to use an experimental design which avoids pretests altogether.

Selection and X Interaction: consider an experiment on teaching where the researcher has been turned down by nine school systems and is finally accepted by the tenth. This tenth almost certainly differs in many ways from the other nine, and from the universe of schools to which we would like to generalize. It is thus nonrepresentative. Its staff may well have higher morale, or less fear of being evaluated, or more zeal for improvement, than does the average school. Here the effects we find, while internally valid, are specific to such schools. To help us judge such matters, research reports should describe the schools and classes which were asked to cooperate but refused, so that the reader can estimate the severity of possible selective biases. Generally speaking, the greater amount of cooperation the experiment requires, the more it disrupts routine, and the higher the refusal rate, the greater will be the selection and *X* interaction.

Selection biases of this sort often arise simply because an experimenter does not bother to allow a more representative selection of schools the opportunity of refusing. Thus much educational research is done in schools with the highest percentage of university professor's children enrolled. While perfect sampling representativeness is impossible to obtain, and is almost totally neglected in many sciences, it both can and should be emphasized in educational research. One way to improve sampling is to reduce the number of students or classrooms participating from a given school or grade and to increase the number of schools and grades in which the experiment is conducted. Obviously we cannot now conduct experiments on samples representatively drawn from all classrooms in the world or even the U.S. We must learn how far we can generalize an internally valid finding only piece by piece, by testing the treatments over a wide variety of conditions.

Reactive Arrangements: whenever an experiment has obviously artificial aspects we need to fear *reactive arrangements*. A major source of unrepresentativeness in most experiments is simply the student's knowledge that he is participating in an "experiment." A problem solving task is automatically generated, in which the procedures and experimental treatment are reacted to not only as simple stimuli, but as clues in divining the experimenter's intent. The play-acting, outguessing, up-for-inspection, or I'm-a-guinea-pig attitudes so caused are not representative of the school setting, seriously hampering generalization of the effect of *X*. Research involving reactive arrangements should not be abandoned. Rather, it should be improved.

Any aspect of the experimental procedure may produce *reactive-arrangements* effects. Pre-testing, for example, may produce such effects; first, by occurring at all, and second, by the test content. The randomization and assignment of individuals to treatments may also cause such problems. Consider the effect upon a classroom when a randomly selected half of the pupils in a class are sent to a separate room. This action, plus the presence of strange "teachers," will certainly create expectations of something unusual, with wonder and active puzzling as to purpose. The presentation of the treatment *X*, if an out-of-ordinary event, could have a similar effect. Presumably, even a posttest could create such attitudes. This becomes more likely, the more obvious the connection between the experimental treatment and the posttest content.

In much educational research there is no need whatsoever for the students to know that an experiment is going on. (It might be nice also to keep the teachers from knowing, but this is not usually feasible.) Several features may make such disguise possible. If the *X*s are variants on usual classroom events occurring at plausible times, then one-third of the battle is won when these treatments occur without special announcement. If the *O*s are similarly imbedded as regular examinations, the second requirement is achieved. If the *X*s are communications to individual students, then randomization can be achieved without the physical transportation of randomly equivalent samples to different classrooms, etc.

As a result of such considerations, and as a result of personal observations of experimenters who have published data in spite of having such poor rapport that their findings were quite misleading, the present authors recommend that experimentation within schools must, whenever possible, be conducted by regular staff of the schools concerned, especially when findings are to be generalized to other classroom situations. At present, there seem to be two main types of "experimentation" going on within schools: (1) research "imposed" upon the school by an outsider, who has his own ax to grind and whose goal is not immediate action (change) by the school; and (2) the so-called "action" researcher, who tries to get teachers themselves to be "experimenters," using that word quite loosely. The first researcher gets results that may be rigorous but not applicable. The latter gets results that may be highly applicable but probably not accurate because of extreme lack of rigor in the research. An alternative model is for the ideas for classroom research to originate with teachers and other school personnel, with designs to test these ideas worked out cooperatively with specialists in research methodology, and then the bulk of the experimentation be carried out by the idea-producers themselves. The appropriate statistical analyses could be done by the research methodologist and the results fed back to the group via a trained intermediary (supervisor, director of research in the school system, etc.) who has served as intermediary all along. Results should then be relevant and "correct."

Multiple-X Interference is a problem whenever multiple treatments are applied to the same persons. Relatively few designs involve such multiple application of treatments. The findings of such designs can usually be generalized only to populations in which the same sequence of events has been established.

In summary, the average student or potential researcher reading the previous discussion probably now has more worries about experimental design than he had when he began. These are good, if they lead to the design and execution of better experiments and to more care in drawing inferences from results. They are bad, however, if they discourage one about experimental control, or lead to less adequate methods of investigation. Another danger is that this formidable list of sources of invalidity might discourage one, especially about the *quasi*-experimental designs, where full experimental control is lacking in the first place. Such an effect would indeed be the opposite of our intention. Every experiment is imperfect.

What such a checklist of validity criteria can do is to make an experimenter more aware of the inescapable imperfections in any design he does pick so that he can be aware of competing interpretations of his data. He should, of course, design the very best experiment possible in the situation. He should also deliberately seek out those artificial and natural laboratories providing the best opportunities for control. But beyond that he should *go ahead with experiment and interpretation*.

Three True Experiments

A "true" experiment is one in which the experimental units (e.g., students) are assigned to treatment groups randomly in some systematic, predetermined way. "Quasi-experimental" designs, on the other hand, are those in which random assignment to treatments is not possible. Still other designs, of even lower validity, may be called "pre-experimental." We shall first consider three true experimental designs.

The three basic designs to be treated in this section are in general the best available. Design 1 is the most used, and therefore receives more attention here. Note that all three designs are presented in terms of a single X compared with *no* X. Multiple X designs, those with more than 2 treatments, are also possible and important. They are discussed later.

1. The Pretest-Posttest Control Group Design

$$\begin{array}{llll} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array}$$

In this design, as in the other true experimental designs to be considered, the equivalence of groups is achieved by randomization, designated above by R.

Let us consider how this design controls for the various threats to validity. *History* is controlled, since events in time which might produce an $O_1 - O_2$ difference also produce an $O_3 - O_4$ difference. Many applications of Design 1, however, (or 2 or 3, for that matter) do not completely control for history. If all the randomly assigned students in the experimental group are treated in one session, for instance, and the control students in another session, then the irrelevant unique events in either session (the obstreperous joke, the fire across the street, the experimenter's introductory remarks, etc.) become rival hypotheses explaining the $\underline{O}_1 - \underline{O}_2$ vs. $\underline{O}_3 - \underline{O}_4$ difference. Since current "best" practice in educational experimentation usually fails to control such events, it may not seem a serious limitation. But unfortunately, our "best" practice is producing experiments which are too often impossible to replicate. And this limitation is serious indeed.

In this first design, Maturation and Testing are controlled, since they are the same in both experimental and control groups. Also, Instrumentation is easily controlled where a fixed instrument such as a printed test is used. Where observers or interviewers are used, however, the problem is more difficult. If observers are too few to be randomly assigned to observations of single sessions, then each observer should be used for both experimental and control sessions. Observers should not be told which students are receiving which treatments, so that this knowledge will not bias their ratings or records. The tape recording of group proceedings may in some cases help, since then judges must evaluate all transcriptions without knowledge of their source (pretest or posttest, experimental or control).

Regression is controlled as far as mean differences are concerned, no matter how extreme the group is on pretest scores, if indeed both experimental and control groups are randomly assigned from this same extreme pool. Where so assigned, the control group will regress as much as the experimental group.

Selection is ruled out as an explanation of the difference, since randomization assures a kind of group equality at time $\underline{R}$.

Mortality refers to lost cases, cases on which only partial data are available. These are troublesome to handle, and are commonly swept under the rug. Typically, educational experiments are spread out over days, weeks, and months. Sometimes the experimental treatments may involve attendance at certain sessions, while the control condition does not. Absences in any of the sessions produce "mortality," which can subtly bias the sample obtained. If we disregard those who fail to show up for experimental sessions, we thereby shrink the experimental group more than the control group, and so bias the experimental group in the direction of the conscientious, for example, or the healthy. While not usually employed, a better procedure would be to use all the selected experimental and control students who completed both pretest and posttest, including those in the experimental group who failed to get the X. This procedure will bother some, since it counts with X those who did not receive X, but it does avoid the sampling bias.

The mortality problem can be seen as greatly exaggerated, but still solvable, in any study using volunteers. For example, suppose a sample of poor readers in a high school is invited to participate in voluntary remedial sessions, while an equivalent group is not invited. Of the invited group, perhaps 30% participate. Posttest scores, like pretest scores, come from standard reading achievement tests administered to all subjects. It is unfair to compare the 30% volunteers with the total control group, because they represent those most disturbed by the pretest, those perhaps most likely to be vigorous in self-improvement, etc. It is impossible to locate the exact counterpart of these students in the control group. While it also may seem unfair to compare posttest scores of the total invited group with the total uninvited, this is a more acceptable, though very conservative solution.

A far better solution still is to assign experimental and control groups randomly from among all seekers of remedial treatment. This can be done fairly by postponing the remediation given the control group until a later date. (It may be found ineffective, anyway.)

2. The Random Four-Group Design

$$\begin{array}{llll} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \\ R & & X & O_5 \\ R & & & O_6 \end{array}$$

While Design 1 is more frequently used, Design 2 deservedly has higher prestige since it overcomes some of the limitations of Design 1. By adding a new experimental and a new control group without pretests, we can now estimate the effects of *testing*, and the *testing and X interaction*. In this way, we increase generalizability, and are able to make more comparisons: ($O_2 > O_1$, $O_2 > O_4$, $O_5 > O_6$, and $O_5 > O_3$). If these comparisons are all consistent, obviously the strength of our inference about X is increased.

3. The Posttest-Only Control Group Design

$$\begin{array}{lll} R & X & O \\ R & & O \end{array}$$

While many research workers in education and psychology lean naturally to pretests, they are *not* actually essential to true experimental designs. Most researchers are reluctant to give up "knowing for sure" that the experimental and control groups were "equal" before the differential experimental treatment. Nevertheless, as we have already said, the most adequate all-purpose procedure for insuring the equivalence of groups is *randomization*. Within the limits of confidence stated by certain tests of significance, randomization *can* suffice without pretest. In educational research, we must often experiment with entirely new subject matter, for which pretests in the ordinary sense are impossible. Design 3 is appropriately unbiased in this situation and, in fact, in all settings where true randomization of treatments is possible. It is therefore the most universally

Table 1

Sources of Invalidity for Designs 1 through 6

	SIX DESIGNS	Sources of Invalidity											
		Internal								External			
		History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
TRUE EXPERIMENTAL	1. Pretest-Posttest Control Group Design R O X O R O O	+	+	+	+	+	+	+	+	-	?	?	
TRUE EXPERIMENTAL	2. Random Four-Group Design R O X O R O O R X O R O	+	+	+	+	+	+	+	+	+	?	?	
TRUE EXPERIMENTAL	3. Posttest-Only Control Group Design R X O R O	+	+	+	+	+	+	+	+	+	?	?	
PRE-EXPERIMENTAL	4. One-Shot Case Study X O	-	-				-	-					
PRE-EXPERIMENTAL	5. One-Group Pretest-Posttest Design O X O	-	-	-	-	?	+	+	-	-	-	?	
PRE-EXPERIMENTAL	6. Static Group Comparison X O O	+	?	+	+	+	-	-	-				

Note: A plus indicates control of a source of invalidity. A minus indicates lack of control. A question mark indicates that the source is controlled in some situations but not in others. A blank indicates that the source is not relevant.

applicable of the true experimental designs, though for small numbers of experimental units it may lack "power" -- that is, not be sensitive enough to small or even moderately large real effects of treatments.

A comparison of the three experimental designs is shown in Table 1.

More Complex True Experiments

Each of the three preceding designs can be extended, by adding other groups; e.g., $\underline{X}_1$, $\underline{X}_2$, $\underline{X}_3$, etc. If a control group is regarded as a treatment group, then in Designs 1 and 3, there would be one group for each treatment. In Design 2, there would be two groups (one pretested, one not) for each treatment, and a two-classification ("two-way") analysis could be performed. Very frequently, two or more treatment variables, each at several "levels," will be employed, giving a series of groups that can be designated $\underline{X}_{a_1}$, $\underline{X}_{b_1}$, $\underline{X}_{a_1}$ $\underline{X}_{b_2}$, $\underline{X}_{a_1}$ $\underline{X}_{b_3}$, ..., $\underline{X}_{a_2}$ $\underline{X}_{b_1}$, etc.

Such elaborations, produce the traumatizing mysteries of factorial design, randomized blocks, split plots, Greco-Latin squares, fractional replication, etc.

The details of statistical analyses for complex designs cannot be presented here. The concept of interaction, however, plays a central part in experimental thinking. Because a detailed knowledge of interaction is highly relevant to the external validity problem, it will be discussed here in some detail.

Interaction. We have already used the term interaction where the untrained reader could understand it. As before, we are concerned with generalizability. In Figure 2, five possible outcomes are shown in graphic form.

In Figure 2a, A has a visible effect ($A_1 < A_2 < A_3$). So does B ($B_1 < B_2 < B_3$). But there is no interaction between A and B. The lines are therefore parallel. All of the other graphs show significant interactions between A and B. That is, the effect of A itself is dependent upon the specific value of B. Such interaction is of great educational and scientific importance. In Figure 2d, for example, A

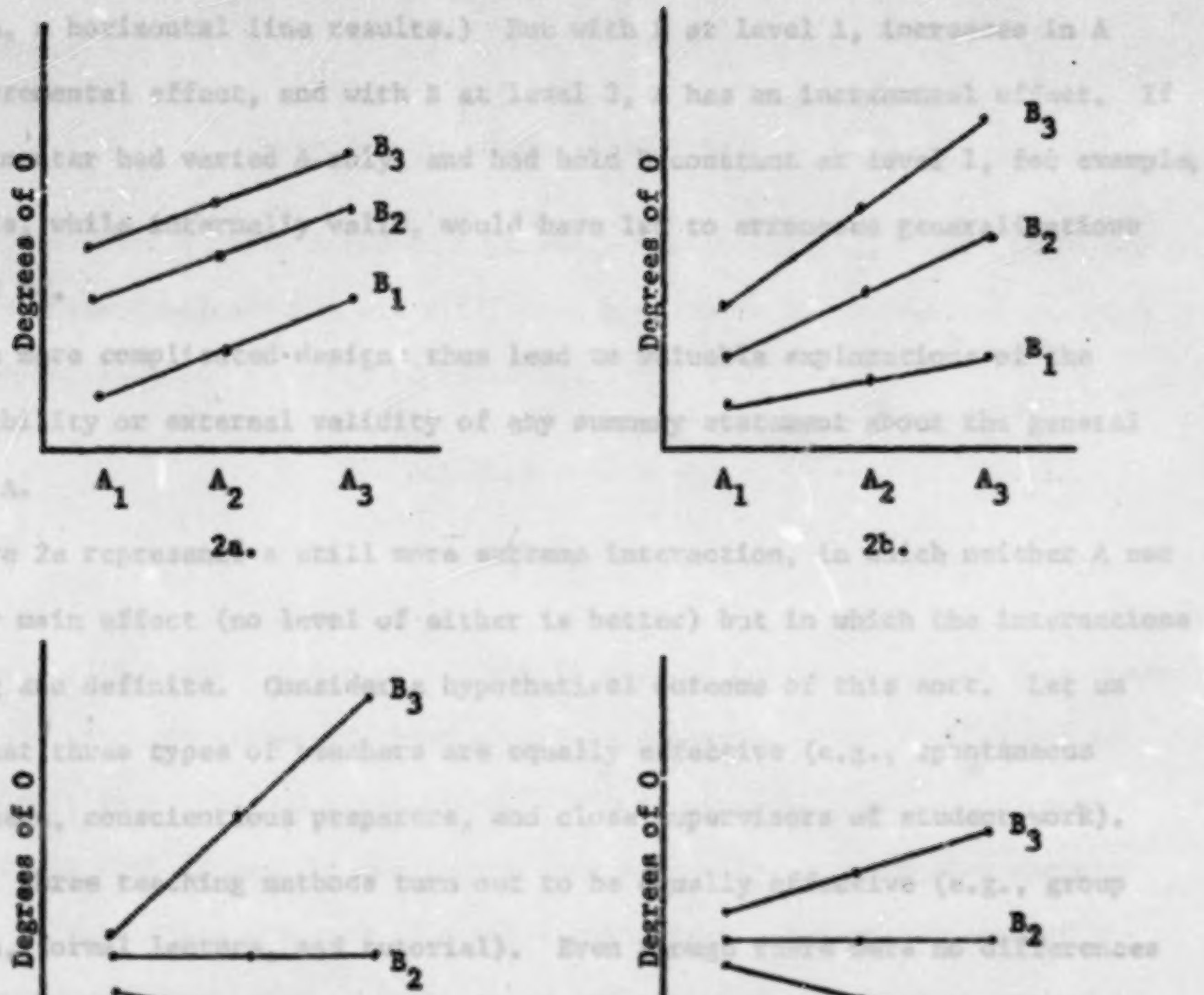

2a. 2b. 2c. 2d. 2e.

Figure 2. Some possible outcomes of a complex experiment.

does not have a main effect (i.e., if one averages the values of all three Bs for each A, a horizontal line results.) But with B at level 1, increases in A have a decremental effect, and with B at level 3, A has an incremental effect. If the experimenter had varied A only, and had held B constant at level 1, for example, the results, while internally valid, would have led to erroneous generalizations for B_2 and B_3.

These more complicated designs thus lead to valuable explorations of the generalizability or external validity of any summary statement about the general effect of A.

Figure 2e represents a still more extreme interaction, in which neither A nor B have any main effect (no level of either is better) but in which the interactions are strong and definite. Consider a hypothetical outcome of this sort. Let us suppose that three types of teachers are equally effective (e.g., spontaneous extemporizers, conscientious preparers, and close supervisors of student work). Similarly, three teaching methods turn out to be equally effective (e.g., group discussion, formal lecture, and tutorial). Even though there were no differences among teacher-type or teaching method, teaching methods could still interact strongly with teacher types. The spontaneous-extemporizer could do best with group discussion and poorest with tutorial; the close supervisor could have the reverse order of effectiveness. Such would be the principal finding of the experiment, and it would be an important one.

It is helpful to distinguish between kinds of interactions. Perhaps the concept, "monotonic interactions," will help us. In 2b, as in 2a, there is a main effect of both A and B, and A has the same directional effect ($A_1 < A_2 < A_3$) for every separate B value. Therefore we can be more confident in stating our general conclusions about the effect of A than we can in case 2c. We might in fact be nearly as confident of the generality of A's main effect over the three levels of B in an experiment like 2b as in the interaction-free 2a. Certainly, before generalizing our results in any experiment we should always plot and examine them in detail. Some "monotonic" or single-directional interactions produce little or no limitations.

Three Pre-Experimental Designs

Having considered three forms of true experimental designs, we are in position to evaluate three designs which are still used frequently in educational research but are greatly inferior in terms of validity.

4. The One-Shot Case Study

X O

Frequently in education a single group is studied only once, after some treatment or event which we assume may have caused some change. Such studies are so lacking control that they are scientifically almost useless. Yet because some researchers use them frequently and regard them seriously, they require comment.

Basic to scientific evidence is careful comparison. Any appearances of absolute or intrinsic knowledge about singular isolated objects, is upon analysis found to be an illusion. Therefore, scientific evidence involves at least one comparison. For such a comparison to be useful, both of its sides should be made with similar care and precision. In the case studies of Design 4, a carefully studied single instance is, in a sense, being compared with some other events casually observed or remembered. The inferences are based upon general beliefs of what behavior might have been without X. Such studies often involve tedious collection of specific detail, careful observation, testing, and the like. In fact, they are guilty of misplaced precision. How much more valuable such a study would be if the one set of observations were reduced in half, with the saved effort being directed to the study in equal detail of an appropriate comparison instance. At the present time it seems almost unethical to allow case studies of this nature (i.e., involving a single group observed only at one time) to be used as thesis or dissertation material in education.

"Standardized" tests in such case studies provide limited help, at best, since sources of difference other than X are so numerous that "national norms" are almost useless as controls. Neither can the case study be justified as a "bench mark" for future case studies. The many uncontrolled sources of difference between a present case study and potential future ones are so numerous that their employment is hopeless. In general, then, it would be better to put half of the descriptive effort into both sides of an interesting comparison than to continue with the design X O.

5. The One-Group Pretest-Posttest Design

$$O_1 \quad X \quad O_2$$

While this design is also widely used in educational research, and is judged enough better than Design 4 to be worth doing where nothing else better can be done, it is presented here as another "bad example." It is infested with weaknesses, as summarized in Table 1.

6. The Static Group Comparison Design

$$\begin{array}{cc} X & O_1 \\ \hdashline & O_2 \end{array}$$

In this design, one intact group which has experienced X is compared with one which has not, to establish the effect of X. Here as elsewhere, the broken line indicates that the groups are intact and that individuals are not assigned randomly to such groups. Examples of this design might be the comparison of school systems requiring the bachelors degree of teachers (the X) versus those not; the comparison of students in classes given speed-reading training versus those not; the comparison of those who heard a certain TV program with those who did not; etc. In marked contrast with the "true" Design 3, here there is no formal means of certifying that the groups would have been equivalent without any experiment at all. While the Static Group Comparison Design is inherently stronger than Designs 4 or 5, it does not approach the validity of a true experiment.

Quasi-Experimental Designs

As stated previously, in a quasi-experimental design there is no random assignment of subjects to treatments. There are many natural social settings in which the researcher can introduce some experimental control into his procedures even though he lacks full control over the assignment to treatments. We should be aware of such possibilities and seize them when the opportunity arises. We shall here survey the strengths and weaknesses of a collection of quasi-experimental designs, each one useful *where better designs are not feasible*. When full experimental control is lacking, however, it is all the more imperative that we be fully aware of the variables not controlled in a particular design.

7. The Time-Series Design

$$\underline{O}_1 \quad \underline{O}_2 \quad \underline{O}_3 \quad \underline{O}_4 \quad \underline{X} \quad \underline{O}_5 \quad \underline{O}_6 \quad \underline{O}_7 \quad \underline{O}_8$$

In the Time-Series Design we periodically measure some group or individual, introduce an experimental change in the series, and then continue measuring periodically. We are interested in how the pre-$\underline{X}$ measures differ from the post-$\underline{X}$ measures. Some possible outcomes of such an experiment are graphed and explained in Figure 3.

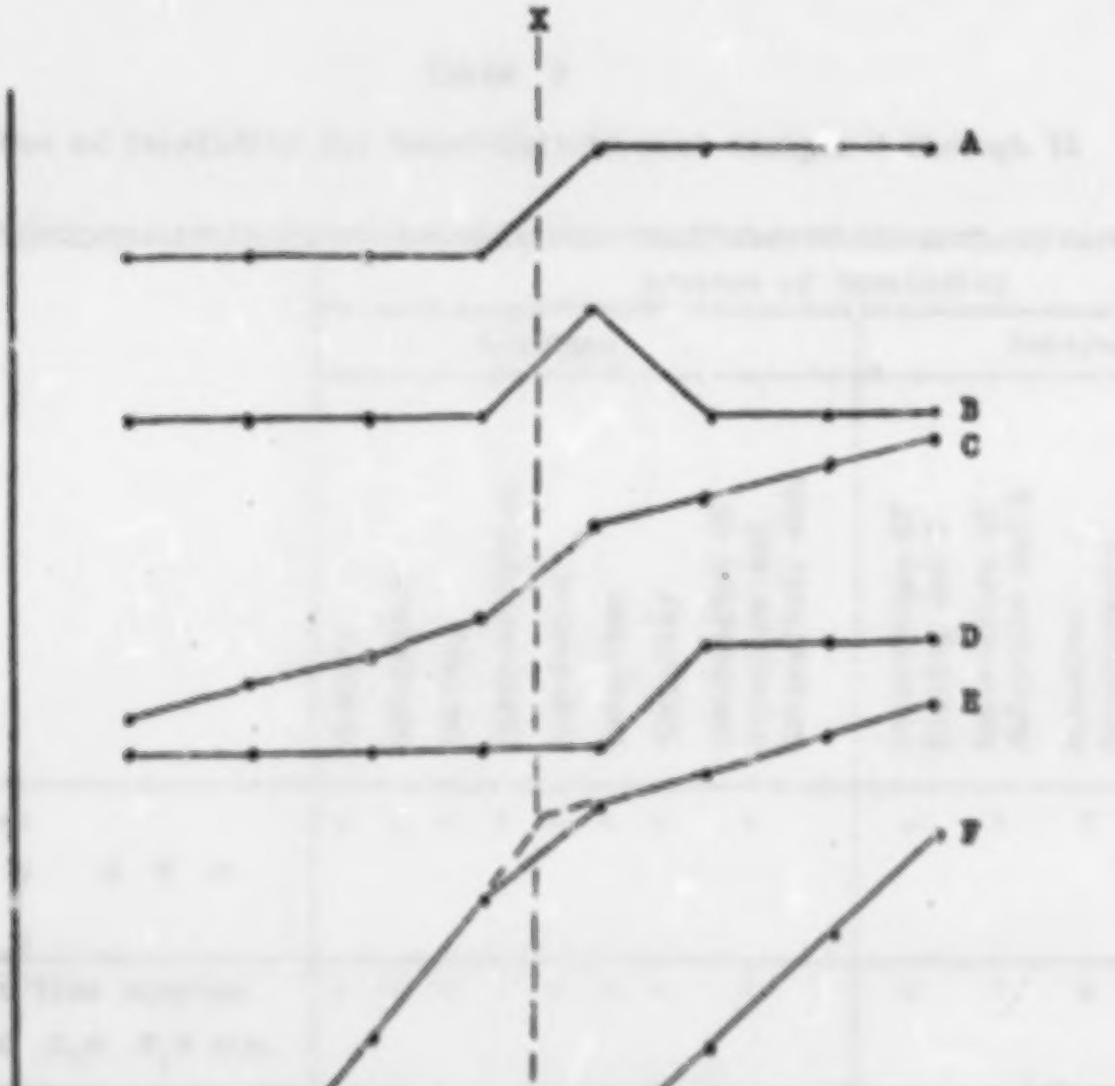


Figure 3 Some possible outcome patterns from the introduction of an experimental variable at point X into a time-series of measurements, O_1 through O_8. Except for D, the $O_5 - O_4$ gain is the same for all time-series, while the legitimacy of inferring an effect varies widely, being strongest in A and B, and totally unjustified in F, G, and H.

Table 2

Sources of Invalidity for Quasi-Experimental Designs 7 through 14

QUASI-EXPERIMENTAL DESIGNS	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
7. Time Series O O O O X O O O O	-	+	+	?	+	+	+	+	-	?	?	
8. Equivalent Time Samples X_1O X_0O X_1O X_0O X_1O etc.	+	+	+	+	+	+	+	+	-	?	-	-
9. Equivalent Materials Samples M_aX_1O M_bX_2O M_cX_1O M_dX_2O M_eX_1O etc.	+	+	+	+	+	+	+	+	-	?	?	-
10. Nonequivalent Control O X O - - - - - - - - - O O	+	+	+	+	?	+	+	-	-	?	?	
11. Separate-Sample Pretest-Posttest R O (X) R X O	-	-	+	?	+	+	-	-	+	+	+	
11a. R O (X) R X O - - - - - - - - - R O (X) R X O - - - - - - - - -	+	-	+	?	+	+	-	+	+	+	+	

Table 2 (Continued)

Sources of Invalidity for Quasi-Experimental Designs 12 through 14

QUASI-EXPERIMENTAL DESIGNS (Continued)	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
12. Separate-Sample Pretest-Posttest Control Group R O (X) R X O - - - - - - - - - - R O R O	+	+	+	+	+	+	+	-	+	+	+	
12a R O (X) R X O - - - - - - - - - - R O (X) R X O - - - - - - - - - - R O (X) R X O (R) R O R O - - - - - - - - - - R O R O - - - - - - - - - - R O R O (R')	+	+	+	+	+	+	+	+	+	+	+	
13. Multiple Time Series O O O X O O O - - - - - - - - - - O O O O O O	+	+	+	+	+	+	+	+	-	-	?	
14. Institutional Cycle Class A X O_1 - - - - - - - - - - Class B O_2 X O_3 - - - - - - - - - - $O_2 < O_1$	+	-	+	+	?	-	?	-	+	?	+	
$O_2 < O_3$	-	-	-	?	?	+	+	-	-	?	+	

8. Equivalent Time-Samples

$$\underline{X_1O} \quad \underline{X_0O} \quad \underline{X_1O} \quad \underline{X_0O}$$

This common one-group experiment uses two equivalent samples of *occasions*. In one of these, the experimental variable is present, and in another it is absent. While the experiment appears to be merely an extension of the Time-Series experiment, it is really more like a two-group experiment since the analysis involves comparisons of groups with and without $\underline{X}$. For most purposes a uniform alternation of treatments or a too-consistent time spacing is undesirable. Any daily, weekly, or monthly cycle may produce influences which will be confused with treatment effects. Consider an experiment by Kerr (1945), who used Design 8 to study effects of music upon industrial production. Each of his experiments involved a single experimental group with a randomized, equivalent sample of days over periods of months. Thus in one experiment he was able to compare 56 music days with 51 days without music, and in another he was able to compare three different types of music, each represented by equivalent samples of 14 days.

As employed by Kerr, Design 8 seems fully internally valid. *History*, the major weakness of the Time-Series experiment, is controlled by presenting $\underline{X}$ on numerous separate occasions, making any rival explanation based on extraneous events extremely unlikely. The reactive effect of arrangements caused by the awareness of experimentation represents a particular weakness of this experiment. Where separate groups are getting separate $\underline{X}$s, it is possible (particularly under Design 3) to have them totally unaware of the presence of an experiment or of the treatments being compared. This is not so when a single group is involved, and when it is repeatedly being exposed to one condition or another, e.g., to one kind

---

Kerr, W. A. Experiments on the effect of music upon factory production. *Appl*. *Psychol*. *Mongr*., 1945, #5.

of music vs. another in Kerr's experiments. As to the interaction of *selection* and *X*: there is the usual limitation. We can generalize our effects only to the particular type of population involved.

This design also carries the multiple-X interference hazard to external validity. This hazard is common to all experiments in which repeated *X*s are presented to the *same* set of persons. That is, the effect of *X* can be generalized only to populations which have experienced repeated *X*s.

Note, however, that the effects of many aspects of teaching are limited to the actual presence of *X*. Where this is true, Design 8 is quite valuable. Suppose a teacher questions the value of oral recitation versus individual silent study. By varying these two procedures over a series of lesson units, one can arrange an interpretable experiment.

The effect of the presence of a parent-observer in the classroom upon students' volunteer discussion could be studied in the same way. Awareness of such designs can place an experimental testing of alternatives within the grasp of the individual teacher.

9. The Equivalent Materials Design

$$\underline{M_a X_1 O} \quad \underline{M_b X_0 O} \quad \underline{M_c X_1 O} \quad \underline{M_d X_0 O} \quad \text{etc.,}$$

where the M's are equivalent materials.

Consider an illustration of this design, a study of conformity to group opinion. Suppose we ask students to express their opinions on a number of issues presented in a long questionnaire. These questions are then divided into two groups as equivalent as possible. At a later time the questionnaires are handed back to the students indicating for each item the group vote. These votes are falsified, to indicate majorities in opposite directions for the two samples of items. As a post-*X* measure, the students are asked to vote again on all items. Depending upon the adequacy of the sampling equivalence of the two sets of items, the differences in shifts between the two experimental treatments provide an experimental demonstration of the effects of the reporting of group opinions, even in the absence of any control group of persons.

Like Design 8, Design 9 has internal validity on all points, and in general for the same reasons. As to external validity, it may be noted that like all experiments involving repeated measures, the effects may be quite specific to persons repeatedly measured.

Reactive arrangements are easier to avoid than in Design 8 because of the heterogeneity of the materials and the greater possibility that the subjects will not be aware they are getting different treatments at different times for different items. This would be so for the conformity study. Interference among the levels of the experimental variable or interference among the materials seem likely to be a definite weakness for this experiment, as for Design 8.

10. The Nonequivalent Control Group Design

$$\begin{array}{ccc} O & X & O \\ \text{- - -} & \text{- - -} & \text{- - -} \\ O & & O \end{array}$$

This is one of the most widespread designs in educational research. The experimental and control groups are naturally intact, with no randomization possible. Yet the groups are assumed "equivalent" and pretests are used to control any differences. The assignment of *X* to one intact group or the other is under the experimenter's control, and is made randomly.

Two things need to be kept clear about this design: First, it should not be confused with Design 1, the Pretest-Posttest Control Group Design, in which experimental subjects are assigned *randomly* from a common population to experimental and control groups. Second, in spite of this, Design 10 should be recognized as well worth using in many instances in which Designs 1, 2, or 3 are virtually impossible. Even if pretest scores are not equivalent, this design is superior to Design 5, the One-Group Pretest-Posttest Design. The more similar the experimental and the control groups, and the more similar the scores on the pretest, the more effective is this control group.

The design thus can sometimes control the main effects of history, maturation, testing, and instrumentation. If these are offered as alternative explanations for a gain specific to the control group, one must assume an interaction between such variables and the other differences between the experimental and control groups. While in general such interactions are unlikely, there are a number of situations in which they might occur. Perhaps most common are interactions involving maturation.

A concrete illustration from research on teaching may make this point clear. Sanford and Hemphill's (1952) study of the effects of a psychology course at

---

Sanford, F. H., & Hemphill, J.K. An evaluation of a brief course in psychology at the U. S. Naval Academy. Educ. Psychol. Measmt., 1952, 12, 194-216.

Annapolis provides an illustration of Design 10. In this study, a Sophomore Class provided the experimental group and a Junior Class the control group. The greater gains for the experimental group might be explained away as a part of some general sophistication process more active in the first two years than in later years. Thus there might be a natural interaction between the class years and other changes (maturation) in these groups, which would be mistaken for the psychology course.

Regression provides the other major internal validity problem for Design 10. As indicated by the "?" in Table 2, this hazard is avoidable, but is more tripped over than avoided. If either comparison group has been selected for its extreme scores on O or some other measure, then a difference in shift from pretest to posttest between the two groups may well be a product of regression rather than the effect of X.

There is a stubborn misleading tradition in educational experimentation that regards matching as an appropriate and sufficient substitute for random assignment to treatment. This error has been accompanied by a failure to distinguish Designs 1 and 10. If in Design 10 the means of the groups are substantially different, then matching not only fails to provide the intended equation but also insures the occurrence of unwanted regression effects; it is certain that the two groups will differ on their posttest scores regardless of X.

There are actually two versions of Design 10, and one is greatly superior to the other. On the one hand, an experimenter may have two natural groups available, e.g., two classrooms, and have free choice in deciding which gets X, or he at least has no reason to suspect a biased assignment of X. Even though the groups may differ in initial means on *O*, such a study may approach true experimentation. On the other hand, the experimenter may have no such assurance. The experimental group may have deliberately sought *X*, with no control group available from this same population of seekers. Here, the assumption of uniform regression between experimental and control groups is less likely. Also, *selection-maturation interaction* (and the other selection interactions) becomes more probable. The "self-selected" design is thus much weaker. Since the requirements of Design 10 are likely to put fewer limitations on our freedom to sample widely than does Design 1, *selection* is usually less a problem than in a true experiment.

The threat to external validity represented by *reactive arrangements* is present, but likewise probably to a lesser degree than in most true experiments, such as Design 1. Where one has the alternative of using two intact classrooms with Design 10, or taking random samples of the students out of the classrooms for different experimental treatments using Design 1, 2, or 3, the latter arrangement is almost certain to be the more reactive, creating more awareness of experiment, "I'm a guinea pig" attitude, and the like.

11. The Separate-Sample Pretest-Posttest Design

R *O* (*X*)
R *X* *O*

For large populations, such as schools, when we cannot randomly separate groups for different experimental treatments, we *can* sometimes exercise experimental control over the *O*. Such control permits Design 11. The diagram above shows two randomly assigned subgroups. Each is exposed to *X*. One sample is measured before *X*, the other is measured after X. The parentheses around the pretested *X* indicate that it is irrelevant to the design; for experimental purpoces we are interested only in the pretest. The design is not a strong one, as is indicated by its row in Table 2. Nevertheless, it may frequently be the only one feasible, and is often worth doing. It is well to note its superiority over the

One-Group Pretest-Posttest Design, Design 5, through its control of both *testing* and the *testing* and *X* interaction.

We may call this design and the remaining designs to be discussed "patched-up," since we add features to control specific factors, piecemeal (in contrast with the neater "true" experiments, in which a single control group controls for all of the threats to internal validity). The main weakness of the design is its failure to control for *history*; extraneous events could account for differences even without *X*.

For pretests and posttests separated in time by several months, *mortality* can be a problem in Design 11. If both samples are selected at the same time (point *R*), then the more time that elapses, the more members of the treatment leave, with the more transient segments of the population being lost, producing a population difference between the comparison groups. In some settings, as in schools, records will enable the elimination of the pretest scores of those who have left by the time of the posttest, thus making the pretest and posttest more comparable.

This design and other "patched-up" designs move the laboratory into the actual situation to which the researcher wishes to generalize, testing the effects of *X* in its natural setting. For this reason, these designs are apt to be superior in *external validity* or generalizability to the "true" experiments of Designs 1, 2, and 3. The "patched-up" designs put so little demand upon the subjects for cooperation, for being at certain places at certain times, etc., that representative sampling from populations specified in advance can be employed. In Designs 11 and 12, (and, to be sure, in some variants on Designs 1 and 3, where *X* and *O* are delivered through individual contacts, etc.) representative sampling is possible. A few question marks are none the less distributed in the sampling-*X* interaction column of Table 2. These question marks indicate that in general practice the units are selected for reasons of cooperativeness and accessability, which make them likely to be atypical of the universe to which one wants to generalize.

12. The Separate-Sample Pretest-Posttest Control Group Design

$$\begin{array}{llll} \underline{R} & \underline{O} & \underline{(X)} & \\ \underline{R} & & \underline{X} & \underline{O} \\ \hdashline \underline{R} & \underline{O} & & \\ \underline{R} & & & \underline{O} \end{array}$$

A still better "patched-up" design is Design 12. It is commonly used when X must be presented to all individuals in some intact groups but can be withheld completely from others. If there are comparable groups from which X can be withheld, then to Design 11 we may add a "posttested-only" control group, and so create Design 12.

The two groups above the broken line make a Design 11 design. The two control groups below the line make additional comparisons possible and lead to increased internal validity.

The design is also quite similar to Design 10, since it involves the comparison of pretest and posttest differences of nonequivalent groups. However, in Design 12 we do not retest the same persons, so that we thus avoid the interaction of testing and X. As with Design 10, the internal validity weakness of Design 12 comes from the possibility of mistaking for an effect of X some trend in the experimental group which is in fact unrelated to the treatment. By increasing the number of the units involved, e.g., schools, and by randomly assigning them to the experimental and control treatments, we remove this invalidity and achieve a true experiment. This design can be designated 12a. Its diagraming is complicated by

$$\left\{\begin{array}{l} R'\left\{\begin{array}{llll} R & O & (X) & \\ \underline{R} & & X & O \\ \hdashline R & O & (X) & \\ \underline{R} & & X & O \\ \hdashline R & O & (X) & \\ R & & X & O \end{array}\right. \\[1em] R'\left\{\begin{array}{llll} R & O & & \\ \underline{R} & & & O \\ \hdashline R & O & & \\ \underline{R} & & & O \\ \hdashline R & O & & \\ \underline{R} & & & O \end{array}\right. \end{array}\right.$$

the two randomizations involved. Subjects are randomly assigned to pretest and posttest samples, within each unit; e.g., within each school. This is coded by the pairs of Rs separated by broken lines. The assignment of units (schools) to test-plus -*X* or test-only groups is also random. This is coded by the R's. Thus the only non-random assignment comes in the selection of the units themselves. This is coded by the broken lines.

As can be seen by the row for 12a in Table 2, this design recieves a perfect score for both internal and external validity, the latter on grounds already discussed for Design 11 with further strength on the *selection-X interaction* problem because of the representation of numerous units, rather than a single one. As far as is known, this excellent design has not been used.

13. The Multiple Time-Series Design

O O O O X O O O O

- - - - - - - - - - - - - - - - -

O O O O   O O O O

In studying a major administrative change by time-series data, a researcher might wisely tend to seek out a similar institution not undergoing the *X*, from which to collect a similar "control" time series (ideally with *X* assigned randomly).

This design contains within it (in the Os on either side of the *X*) Design 10, the Nonequivalent Control Group Design, but gains in certainty of interpretation from the multiple measures plotted, since the experimental effect is in a sense twice demonstrated, once against the control, and once against the pre-*X* values in its own series, as in Design 7. In addition, the *selection-maturation interaction* is controlled to the extent that, were the experimental group to show a greater rate of gain, it would appear in the pre-*X* *O*s. This superiority is not too evident from the summary evaluations in Table 2. It is present in the final internal validity column, *interaction of selection and maturation*. Because maturation is controlled for in both experimental and control series, for the reasons discussed

earlier in connection with the Time-Series Design 7, the interaction of selection with maturation, instrumentation, or regression is controlled. An interaction of selection with history, however, still remains a possibility.

As with the Time-Series Design 7, a minus has been entered on the external validity topic of testing-$\underline{X}$ interaction, although as with Design 7, the design can often be used where the testing is nonreactive. The standard precaution about the possible specificity of a demonstrated effect of $\underline{X}$ to the population under study is also recorded in Table 3.

14. The Recurrent Institutional Cycle Design

Class A $\underline{X}$ $\underline{O}_1$

- - - - - - - - - - - - -

Class B $\underline{O}_2$ $\underline{X}$ $\underline{O}_3$

This design is limited to a narrow set of questions and settings, but it opportunistically exploits the features of these settings. The basic strategy involved can be seen by examining the second and third rows of Table 1. It is evident that the patterns of plus and minus marks for Designs 5 and 6 are for the most part complementary, and that right combination of the two might have considerable strength. The design is appropriate in a situation like a school where instruction is continually being presented to a new group of students on a cyclical schedule. The Recurrent Institutional Cycle Design often offers the best design available for investigating the effects of a global instructional program.

The design combines the "longitudinal" and "cross-sectional" approaches commonly employed in developmental research. It involves measuring at the same time a group which has been exposed to $\underline{X}$ and a group which is just about to be exposed to it. This comparison between $\underline{O}_1$ and $\underline{O}_2$ corresponds to the Static Group Comparison, Design 6. Remeasuring the personnel of Class B one cycle later provides the One Group Pretest-Posttest Design 5. In Table 2, on page 24 the first two rows dealing with Design 14 show an analysis of these comparisons. The

cross-sectional comparison of $O_1 > O_2$ provides differences which could not be explained by the effects of history or a test-retest effect. The differences obtained could, however, be due to differences in recruitment from year to year (as indicated by the minus opposite *selection*) or by the fact that the respondents were one year older (the minus for *maturation*). Where the testing is all done at the same time period, shifts in the nature of the measuring instrument or instrumentation seem unlikely.

In the typical comparison of the differences in attitudes of freshmen and sophomores, the effect of *mortality* is also a rival explanation: O_1 and O_2 might differ due to the kind of people that have dropped out from Class A but are still present in Class B. This weakness can be avoided if the experimenter postpones his analysis of the data until Class B has completed its exposure to X and then eliminates from O_2 all of those measures belonging to individuals who later failed to complete the training. The frequent absence of this procedure justifies the insertion of a question mark for mortality.

The *regression* column has two question marks to warn of the possibility of spurious effects if the measure being used in the experimental design is the one on which the acceptance and rejection of candidates for the training course was based. Under these circumstances consistent differences not to be attributed to the effects of X would be anticipated. The pretest-posttest comparison involved in O_2 and O_3, if it provides the same type of difference as does the $O_2 - O_1$ comparison, rules out the rival possibility that differences are due to a shift in the selection or recruitment between the two classes, and also the possibility of mortality as an explanation. Used alone, however, the $O_2 - O_3$ comparison is vulnerable to the rival explanations of history and testing.

The design fails to control for maturation. The seriousness of this limitation will vary depending upon the subject material under investigation. If the experiment deals with the acquisition of a highly esoteric skill or competence, the rival hypothesis of maturation, that just growing older or more experienced in normal everyday societal ways would have produced this gain, may seem highly unlikely.

In the diagram of Design 14 as presented, it is assumed that it will be feasible to present the posttest for one group at the same chronological time as the pretest for another. This is not always the case in situations where we might want to use this design. The following is probably a more accurate portrayal of the typical opportunity in the school situation:

$$\begin{array}{llllll} \text{Class A} & X & O_1 & & & \\ \hdashline \text{Class B} & & & O_2 & X & O_3 \end{array}$$

Such a design lacks the clear-cut control on history in the $O_1 > O_2$. However, the explanation in terms of history could hardly be employed if both comparisons show the effect, except by postulating quite a complicated series of coincidences. Note that any general historical trend, such as we certainly do find with social attitudes, is not confounded with clear-cut experimental results. Such a trend would make O_2 intermediate between O_1 and O_3, while the hypothesis that X has an effect requires O_1 and O_3 to be equal, and O_2 to differ from both in the same direction. In general, with replication of the experiment on several occasions, *history* is unlikely to be a problem even in this version of the design. But, for institutional cycles of less than a calendar year, there may be the possibility of confounding with seasonal variations in attitudes, morale, optimism, intelligence, or what have you. If the X is a course given only in the fall semester, and if between September and January people generally increase in hostility and pessimism due to seasonal climactic factors, this recurrent seasonal trend is confounded with the effects of X in all of its manifestations. For such settings, Designs 10 and 12 are available and to be preferred.

Design 14 in all of its forms must be judged susceptible of a *selection-maturation interaction*, and Table 2 has been marked to indicate this. The seriousness of this limitation depends upon the uniqueness of the population under study.

The ratings on external validity criteria in general follow the pattern of the previous designs containing the same fragments. The question marks in the "Interaction of selection and X" column merely warn that the findings are limited to the institutional cycle under study. Since the X in this design is usually complex, the investigation is apt to be for practical rather than theoretical purposes. If so, it is probably to one institution that we want to generalize.

Correlational and Ex Post Facto Designs

One dimension of "quasi-ness" which has been increasing in the course of the last eight designs is the extent to which X can be manipulated by the experimenter, i.e., intruded into the normal course of events. Certainly the more this is possible, the closer to true experimentation. Designs 7, 10, 11, 12 (but not 12a), and 13 are applicable both for naturally occurring Xs and for Xs deliberately introduced by the experimenter, although the designs are more suspect where X is not under control. Design 14, is, of course, completely limited to a naturally occurring X, and the designs of the present section are still more fully imbedded in the natural setting.

Correlation and Causation

Design 6 is a correlational design of a very weak form. It involves a comparison of but two natural units, differing not only in the presence and absence of X, but also in innumerable other attributes. Each of these other attributes could create differences in the Os, and each therefore provides a plausible rival hypothesis to the hypothesis that X had an effect. We are left with a general rule that the differences between two natural objects are uninterpretable.

Suppose that the comparison is expanded so that we have numerous independent natural instances of X, numerous ones of no-X, and concomitant differences in O. What is the rule now? In so far as the natural instances of X vary among each other in their other attributes, the other attributes become less plausible as rival hypotheses. Correlations of a fairly impressive nature may thus be established, as between heavy smoking and lung cancer. What is the status of such data as evidence of causation analogous to that provided by experiment?

A positive point may first be made: Such data are relevant to causal hypotheses, inasmuch as they expose them to disconfirmation. If a zero correlation is obtained, the credibility of the hypothesis is lessened. If a high correlation occurs, the credibility of the hypothesis is strengthened, in that it has survived a chance of disconfirmation. To put the matter another way, correlation does not necessarily indicate causation, but a causal law of the type producing mean differences in experiments does imply correlation. In any experiment where X has increased O, a positive correlation between presence-absence of X and either posttest scores or gain scores will be found. The absence of such a correlation can rule out many simple, general, causal hypotheses concerning X. In this sense, the relatively inexpensive correlational approach can provide a preliminary survey of hypotheses, and those which survive this can then be checked through the more expensive experimental manipulation.

A perusal of the research in education soon convinces one that the causal interpretation of correlational data is overdone, rather than underdone. Where teacher's behavior and students' behavior are correlated, for example, our cultural stereotypes are such that we almost never consider the possibility of the students' behavior causing the teacher's.

Neither is an inherent temporal order a sufficient demonstration of causality, since selective retention can create a causality in the reverse direction. Consider, for example, possible findings that superintendents with better schools are better educated, or that schools with frequent changes in superintendents have low morale. Almost inevitably we draw the implication that the educational level of superintendents and stable leadership cause better schools. The causal chain could be quite the reverse: better schools (for whatever reasons better) might cause well educated men to stay on, while poorer schools might lead the better educated men to be tempted away into other jobs. Likewise, better schools might well cause superintendents to stay in office longer.

Still more common than misleading reverse correlation is misleading third-variable correlation, in which the lawful determiners of who gets exposed to X also produce high O scores, even without the presence of X. To these instances we will return in the section on the ex post facto design.

The true experiment differs from the correlational study just because the process of randomization disrupts any lawful relationships between the character or antecedents of the students and their exposure to X. Where we have pretests and clear-cut determination of who were exposed and who were not are available, then Designs 10 and 13 may be convincing even without the randomization. But for a design lacking a pretest (imitating Design 3) to occur naturally requires very special circumstances, which almost never happen. Even so, one should keep his eyes open for settings where an exposure to X was lawless, arbitrary, and uncorrelated with prior conditions. Ideally these arbitrary exposure decisions will also be numerous and mutually independent. Furthermore, they should be buttressed by whatever additional evidence is available, no matter how weak, as in the retrospective pretest discussed below.

One such correlational study is of such admirable opportunism as to deserve note here. Barch, Trumbo, and Nangle (1957) used the presence or absence of turn signaling on the part of the car ahead as X, the presence or absence of turn signaling by the following car as O, demonstrating a significant imitation, modeling, or conformity effect, in agreement with the scores of laboratory studies. Lacking any pretest, a causal interpretation of the data is dependent upon the assumption of no relationship between the signaling tendencies of the two cars other than the influence created by the behavior of the lead car. As published, the data seem compelling.

Barch, A.M., Trumbo, D., & Nangle, J. Social setting and conformity to a legal requirement. *J. abnorm. soc. Psychol.*, 1957, 55, 396-397.

Note, however, that any third variables which might affect the signaling frequency of both pairs of drivers in a similar fashion become plausible rival hypotheses. Thus if weather, degree of visability, sex and purpose of the driver as affected by time of day, presence of a parked police car, etc., have effects on both, and if data are pooled across conditions hetergeneous in such third variables, the correlation can be explained without assuming any effect of the lead car's signaling per se.

The Retrospective Pretest

In many military settings in wartime, it is plausible that the differing assignments among men of a common rank and specialty are made through chaotic processes, with negligible regard to special privileges, preferences, or capabilities. Therefore, a comparison of the attitudes of whites who happened to be assigned to racially mixed versus all-white combat infantry units can become of interest for its causal implications (Information and Education Division, 1947).

Information and Education Division, U. S. War Department. Opinions about Negro infantry platoons in white companies of seven divisions. In T. M. Newcomb and E. L. Hartley (Eds), Readings in social psychology. New York: Holt, 1947.

We certainly should not turn our back on such data, but rather should so use them as maximally to rule out plausible rival hypotheses, keeping aware of the remaining sources of invalidity. In this instance, the "posttest" interview not only contained information about present attitudes toward Negroes (those in mixed companies being more favorable) but also asked for the recall of attitudes prior to the present assignment. These retrospective attitudes showed no difference between the two groups, thus increasing the plausibility that prior to the assignment there had been no difference.

Note that the probable direction of memory bias is to distort the past attitudes into agreement with the present ones, or into agreement with what the person

has come to believe to be socially desirable attitudes. Thus memory bias seems more likely to disguise, rather than masquerade as, a significant effect of *X* in these instances.

If studies continue to be made comparing freshman and senior attitudes to show the impact of a college, the use of retrospective pretests to support the other comparisons would seem desirable, as partial curbs to the rival hypotheses of history, selective mortality, and shifts in initial selection. (This is not to endorse any further repetition of such cross-sectional studies, when by now what we need are more longitudinal studies which provide repeated measures over the four-year period, supplemented by repeated cross-sectional surveys in the general manner of a four-year extension of Design 14.)

Ex Post Facto Analyses

The phrase "ex post facto experiment" has come to refer to efforts to simulate experimentation by attempting to match subjects in a Design 6 study on pre-*X* attributes.

In a typical ex post facto study (Chapin, 1955) the *X* was high school and the *O*s were success and community adjustment ten years later, as judged from information obtained in individual interviews. The matching in this study was done from records retained in the high school files (although in similar, still weaker, studies the pre-*X* facts are obtained in post-*X* interviews). Initially the data showed those completing high school to have been more successful. However, they also had higher marks in grammar school, higher parental occupations, younger ages, better neighborhoods, etc. These antecedents might have caused both completion of high school and later success. Did the schooling have any additional effect over and above the head start provided by these background factors? Chapin's "solution" to this question was to examine subsets of students matched on all these background

---

Chapin, F. S. *Experimental designs in sociological research*. New York: Harper, 1947. Revised Edition, 1955.

factors but differing in completion of high school. The addition of each matching factor reduced in turn the posttest discrepancy between the X and no-X groups, but when all matching was done, a significant difference remained. Chapin concluded, although cautiously, that education had an effect. In this study the matching shrank the usable cases from 1194 to 46 - 23 graduates and 23 nongraduates - less than 4% of those originally interviewed. Chapin well argues that 46 comparable cases are better than 1194 non-comparable ones. The tragedy is that his 46 cases are still not comparable, and furthermore, even within his faulty argument, the shrinkage was unnecessary.

He seriously undermatched for two distinct reasons. His first source of undermatching is that matching is subject to differential regression, which would certainly produce in this case a final difference in the direction obtained (as pointed out in discussing matching in connection with Design 10). Regression after matching is certain here because the differences in the matching factors for those successful versus unsuccessful are in the same direction for each factor as the differences between those completing versus not completing high school. Every determinant of exposure to X is likewise, even without X, a determinant of O. All matching variables correlate with X and O in the same direction. While this might not be so of every variable in all ex post facto studies, it is in most if not all published examples. This error, however, and the reduction in number of cases, are avoidable through the modern statistical procedure which supplanted the matching-error in Design 10 (analysis of covariance). There is, however, a second and essentially uncorrectable source of undermatching in Chapin's setting. Greenwood (1945) refers

---

Greenwood, E. Experimental sociology: a study in method. New York: King's Crown Press, 1945.

---

to it as the fact of self-selection of exposure or nonexposure. Exposure is a lawful product of numerous antecedents. In the case of dropping out of high school before completion, we know that there are innumerable determinants beyond the six upon which matching was done. Most of these will likely have a direct effect upon later success, independently of their effect through X. This insures that there will be undermatching over and above the matching-regression effect. Even with the pre-X-predictor and O covariance analysis, a significant treatment effect is interpretable only when all of the jointly contributing matching variables have been included.

Summary

In concluding the discussion of design, a final word of caution is needed about using Tables 1 and 2. These tables have added a degree of order to the discussion as a recurrent outline, and have made it possible for us to be less repetitious than we would otherwise have been. But the placing of specific plusses, minuses, and question marks has been continually equivocal, and is usually an inadequate summary of the corresponding discussion. Moreover, for any specific execution of a design, the checkoff row will probably be somewhat different from the corresponding row in the table.

The tables are best used as an outline for surveying the specific details of an experiment while planning it. In a similar vein, the discussion must not substitute a dogma of the 11 acceptable designs for an earlier dogma of the one or the two acceptable. Rather, it should encourage an open-minded and exploratory orientation to new data collection arrangements, and new consideration of the weaknesses of routine uses of the traditional ones.

APPENDIX B

Form NDEA-VII-1

Budget Bureau No. 51-R281
Approval expires February 28, 1959

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education
Washington 25, D.C.

INSTRUCTIONS FOR PREPARATION OF APPLICATIONS FOR GRANTS TO SUPPORT RESEARCH PROJECTS UNDER THE PROVISIONS OF TITLE VII OF THE NATIONAL DEFENSE EDUCATION ACT OF 1958

Under the provisions of Title VII of Public Law 85-864, 85th Congress, grants-in-aid may be made to public or private non-profit agencies, organizations, and individuals for research and experimentation in more effective utilization of television, radio, motion pictures, and related media for educational purposes. Such research and experimentation must be of value to State or local agencies in the operation of public elementary or secondary schools, and to institutions of higher learning, and may include the development of more effective techniques and methods--(1) for utilizing and adapting motion pictures, video tapes, and other audio-visual aids, filmstrips, slides and other visual aids, recordings (including magnetic tapes) and other auditory aids, and radio or television program scripts for such purposes; (2) for training teachers to utilize such media with maximum effectiveness; and (3) for presenting academic subject matter through such media.

General Instructions

1. Method of applying. The first page of the application must show the following information in the order indicated:

APPLICATION TO THE COMMISSIONER OF EDUCATION, U. S. OFFICE OF EDUCATION, DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, FOR A GRANT TO SUPPORT A RESEARCH PROJECT UNDER THE PROVISIONS OF TITLE VII OF THE NATIONAL DEFENSE EDUCATION ACT OF 1958 (P. L. 85-864)

Submitted by: (Name of institution or agency)

Address: (Address of institution or agency)

Telephone number: (Telephone number of institution or agency)

Initiated by: (Full name and position of the individual who initiated the project, ordinarily the director or chief investigator)

Transmitted by: (Full name and position of the official who is approving the transmittal of the application. This must be someone in a position to assume such authority and to commit the institution or agency to the undertakings projected in the proposal)

Date: (Date of approval)

In addition to the typed names, the proposal should be signed by both individuals on the original or master copy.

2. Mailing address for applications: Send all applications directly to:

Program Consultant
New Educational Media Branch
Office of Education
U. S. Department of Health, Education, and Welfare
Washington 25, D. C.

3. Length of application. Normally an application runs from six to eight pages in length. It should not exceed 12 pages except in exceptional cases.

4. Number of copies required. The application should be typed or otherwise reproduced on white paper of standard size on one side of the sheet. Since all applications are submitted to the Advisory Committee on New Educational Media for review, 30 copies are required.

5. Suggestions. It is desirable that the applicant submit an application as soon as it is prepared in order to allow as much time as possible for processing. There is no need for the application to be accompanied by covering letters or other communications endorsing the project.

Outline of Content

Follow the outline given below in describing the proposed project so that the Advisory Committee may have adequate information upon which to base an objective evaluation. In preparing the description, identify each section by using the number and the underlined words in each item given below:

I. Project title. The title should be concise, descriptive, and specific as possible. Avoid technical terms that are obscure.

II. Problem. Give a brief statement of the problem. Explain its importance in the field of education and explain the significance of the proposed research. Briefly review other related research, citing publications, and suggesting ways in which the study proposed will advance knowledge in the field.

III. Objectives. State the specific questions to be answered or the hypotheses to be tested.

IV. Procedure. Describe clearly the proposed plan of attack. (The following information will not apply in those cases where the proposed research does not fall within the category of controlled experimentation involving standard statistical procedures.)

a. General method and/or experimental design.

1. Controls. If one or more control groups are involved, indicate their origin, composition, and purpose.

2. Sampling plan and procedure. Describe the population on which the study is to be based and tell how the sample is to be taken.

b. Data--types to be gathered and methods to be used. Describe questionnaires, interview guides, tests, and inventories. Describe unpublished research instruments in some detail.

c. Methods of statistical or other analysis.

d. Approximate time schedule.

e. Expected end-product.

f. Publication plans.

V. Personnel. Give the name, title, and a brief statement of the research experience of the principal investigator and other key personnel.

VI. Facilities. Indicate briefly the special facilities and other similar advantages available at your institution or agency for the project as proposed.

VII. Duration. Estimate the total time required for this project as follows:

Total amount of time required: xx years; xx months

Beginning: (Date) Ending: (Date)

Budget and Financing

Start this section on a new page. Amounts should be rounded to the nearest dollar.

VIII. Budget. Make a tabular presentation showing the expenditures which will be involved in the project throughout its entire period. The sum of these itemized costs will be the total cost of the project for all years. Then indicate the amount to be contributed by the institution or agency, and that available from other sources, if any. This will yield the "Total non-Federal support of project." The difference between this and the "Total cost (all years") will constitute the amount designated as "Federal support requested."

Budget

Category		Cost
Personnel (List all position titles, approximate per annum salaries, whether full or part-time, length of employment)		$xxxxxxxx
Services (Itemize such expenses as statistical services, etc.)		xxxxxxxx
Supplies and materials		xxxxxxxx
Other (Specify essential travel, communications, etc.)		xxxxxxxx
Total cost (all years)		$xxxxxxxx
Funds to be supplied by		
Institution or agency	$xxxxxxxx	
Other sources (Specify)	xxxxxxxx	
Total non-Federal support of project		$xxxxxxxx
Federal support requested (Total cost less non-Federal support)		$xxxxxxxx

IX. Other support. (a) In a separate paragraph, indicate whether this research proposal has been submitted to any other agency or organization. (b) Is this an extension of or addition to a previous project supported by the Office of Education?

X. Estimated cost to Federal government by fiscal year. Expenditures are not usually made at a constant rate throughout the course of a project. Therefore, make a tabular presentation similar to the table below which will show estimated distribution of the Federal funds requested by fiscal year. A Federal fiscal year begins on July 1, and continues through June 30 of the following year, which is the year for which the fiscal year is named. Thus, fiscal year 1959 runs from July 1, 1958 through June 30, 1959.

Federal Funds Requested

Category	Federal funds requested
Fiscal year 1959	$xxxxxxxx
Fiscal year 1960	xxxxxxxx
Etc.	xxxxxxxx
Total cost (all years)	$xxxxxxxx

Abstract

XI. Prepare a brief abstract (approximately two pages, doubled spaced) in the following form. This is to be separate from the proposal and submitted in 30 copies.

APPLICATION TO THE COMMISSIONER OF EDUCATION, U. S. OFFICE OF EDUCATION, DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, FOR A GRANT TO SUPPORT A RESEARCH PROJECT UNDER THE PROVISIONS OF TITLE VII OF THE NATIONAL DEFENSE EDUCATION ACT OF 1958 (P. L. 85-864)

Submitted by: (Name of institution or agency)

Principal investigators: (Name and position)

Title: (Title of proposal)

Problem: (Brief statement)

Objectives: (Briefly stated)

Procedure: (Brief description of general method, controls, sampling, statistical analysis, etc., as appropriate)

Time schedule: (Approximate beginning and ending dates)

Budget: Total cost of project .... $xxxxxxxx
Total Federal funds requested .... $xxxxxxxx

APPENDIX C

DIVISION OF BIOSTATISTICS
DEPARTMENT OF PREVENTIVE MEDICINE & PUBLIC HEALTH
SCHOOL OF MEDICINE
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

DESCRIPTION OF PROGRAM # 012

ANALYSIS OF VARIANCE NO. 2

1. General Description:

a. This program is for the IBM 709 to perform an analysis of variance for a complete factorial design.

b. As many as 14 variables may be processed.

c. There is no limitation on the number of replicates.

d. The number of categories or levels of any one variable cannot exceed 999, and input data in one replicate cannot exceed 20,000, namely:

$$L_1 \times L_2 \times \ldots \times L_n \leq 20,000$$

Where $L_1, L_2, \ldots, L_n$ are the numbers of levels of the n variables, and $n \leq 14$.

e. The program can perform a transformation of the input data, if desired, according to the code specified in the Control Card.

f. The number of interactions computed by the program may be limited by an entry on the Control Card.

g. A sample output of this program is included.

2. Cards Preparation:

a. I.D. Card

Col.	1	*punch must be present
Col.	3-7	Job number as approved by WDPC
Col.	9-11	Maximum running time in minutes
Col.	13-15	Maximum pages of printed output
Col.	17-80	Your name

Note: (1) For the estimation of running time, use the following rule of thumb: 2 minutes plus N·R+(T/50) seconds for each problem, where N is the number of variables, R is the number of replicates, and T is the input data in one replicate as explained under 1d. above.

(2) For the estimation of pages of printing, use the following rule of thumb: 5 pages for each problem.

b. Control Card (One control card for each problem)

Col.	1,2	Problem number
Col.	3-8	Scale factor (Sign in the first position with 5 digit scale factor.) See the Input Data Cards section.
Col.	9-12	Number of replicates
Col.	13,14	Number of variables
Col.	15,16	Type of transformation, according to code: 00 no transformation 01 $\sqrt{x}$ transformation 02 $\sqrt{x}+\sqrt{x+1}$ transformation 03 $\log_{10}x$ transformation 04 $\log_{10}(x+c)$ transformation when this transformation is desired, the value of c must be specified in Col. 17-21 of the Control Card. 05 $\sin^{-1}\sqrt{x}$ transformation 06 $\sin^{-1}\sqrt{\frac{x}{n+1}}+\sin^{-1}\sqrt{\frac{x+1}{n+1}}$ transformation
Col.	17-21	value of a constant (c) for $\log_{10}(x+c)$ transformation in the form xx.xx. If it is not used, leave these columns blank.
Col.	22-26	Blank
Col.	27,28	Highest order interaction to be computed by program. (If 0 or the number of variables is entered, all interactions are computed.)
Col.	29-31	Number of levels of 1st variable
Col.	32-34	Number of levels of 2nd variable
Col. etc.	35-37	Number of levels of 3rd variable

c. Input Data Cards:

Since the size of input data can not exceed 5 digits (or 6 if the sign is omitted), it may be necessary to scale the data in order not to lose significant digits in data inputs. The data are punched without decimal points. The locations of the decimal points are determined by the scale factor.

Example:	Input Data	Data Keypunched	Scale Factor
1)	312	+00312	+00000
2)	-31245000	-31245	+00003
3)	.00031245	+31245	-00008

A proper scale factor which applies to all input data must be specified in Col. 3-8 of the Control Card.

c. **Input Data Cards (cont.)**

Consider the 3-variable analysis of variance design, where one variable has 2 levels, the other 2 variables have 3 levels and the entire experiment is replicated twice. The data may be represented in the form

x_{rijk} $r = 1,2$ replicate
$i = 1,2,3$
$j = 1,2,3$
$k = 1,2$.

In preparing data for analysis each replicate of the design must start on a new card. Within each replicate the numbering of the variables by the program proceeds in the following manner:

If the data are keypunched in the order

x_{r111}
x_{r112}
x_{r121}
x_{r122}
x_{r131}
x_{r132}
x_{r211}
.
.
.
x_{r332}

then the program considers the $i = 1,2,3$ variable as variable 1
$j = 1,2,3$ variable as variable 2
$k = 1,2$ variable as variable 3.

This ordering must be maintained within each replicate.

The preparation of Input Data Cards is as follows:

1st card, 1st replicate,	Col.	1-6	a sign and x_{111}
	Col.	7-12	a sign and x_{112}
	Col.	13-18	a sign and x_{121}
	.		.
	.		.
	.		.
	Col.	67-72	a sign and x_{232}
2nd card, 1st replicate,	Col.	1-6	a sign and x_{311}
	Col.	13-18	a sign and x_{312}
	.		.
	.		.
	Col.	31-36	a sign and x_{332}
	Col.	37-72	Blank .

c. Input Data Cards (cont.)

3rd card, 2nd replicate,	Col.	1-6	a sign and x_{111}
		.	.
		.	.
		.	.
	Col.	67-72	a sign and x_{232}
4th card, 2nd replicate,	Col.	1-6	a sign and x_{311}
		.	.
		.	.
		.	.
	Col.	31-36	a sign and x_{332}
	Col.	37-72	Blank

Leading zeros may be indicated by blanks or by punched zeros.

d. End Card:

This card will notify the program of the end of the entire job. This card has the following format

Col. 1-10 all 9 punch

3. Order of Cards:

More than one problem can be processed consecutively. The following example illustrates the order of cards.

1) I.D. Card
2) Program Cards**
3) Control Card for problem 1
4) Input Data Cards

5) Control Card for problem 2
6) Input Data Cards
.
.
.

7) Control Card for the last problem
8) Input Data Cards
9) End Card

** This is a prepared binary deck of 115 cards labelled BIMED # 012.

CARDS FOR PROBLEM

```
*00451 003 003 BIMED 012                                                       I.D.
PROGRAM CARDS-BINARY DECK
01-000020002030000000   00003002002                                           CONTROL
  873  950   92  506  976  957  222  608  1013  1038   100   632DATA 1
  715  922  112  566  947  991  166  633   920  1029   169   618DATA 2
9999999999                                                                     END
```

```
ANALYSIS OF VARIANCE
PROBLEM NO.  1


NUMBER OF VARIABLES     3
NUMBER OF REPLICATES    2
HIGHEST ORDER INTERACTIONS COMPUTED     3

VARIABLE     NO. OF LEVELS
    1              3
    2              2
    3              2

TRANSFORMATION CODE  0


GRAND MEAN              6.56458


SOURCE OF          DEGREES OF               SUMS OF            MEAN
VARIATION            FREEDOM                SQUARES            SQUARES

1                       2                   4.98811            2.49405
2                       1                 198.77773          198.77773
3                       1                  41.21261           41.21261
12                      2                   0.61956            0.30978
13                      2                   0.55256            0.27628
23                      1                  21.26284           21.26284
REMAINDER               2                   0.53777            0.26888
WITHIN REPLICATES      12                   2.45965            0.20497
TOTAL                  23                 270.41084
```

APPENDIX D

DIVISION OF BIOSTATISTICS
DEPARTMENT OF PREVENTIVE MEDICINE AND PUBLIC HEALTH
SCHOOL OF MEDICINE
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

DESCRIPTION OF PROGRAM NO. 013

ANALYSIS OF COVARIANCE

1. **General Description:**

   a. This program is for the IBM 709 to perform analysis of covariance.

   b. Two to six analysis of variance variables may be processed with as many as eight covariates.

   c. As many as 999 replicates are allowed.

   d. The number of categories or levels of any one variable cannot exceed 999, and input data in one replicate cannot exceed the following limitation:

   $$L_1 \times L_2 \times L_3 \times .. \times L_n \leq 2000$$

   Where L_1, L_2, L_3, .., L_n are the numbers of levels of the n variables and $n \leq 6$.

   e. The program can perform a different transformation of variate or any covariate, if desired, according to the code specified in the Transformation Card.

   f. A sample output of this program is shown in the Appendix.

2. **Cards Preparation:**

   a. I. D. Card:

Col.	1	*punch must be present
Col.	3-7	Job number as approved by the WDPC
Col.	9-11	Maximum running time in minutes
Col.	13-15	Maximum pages of printed output
Col.	17-80	Your name

   Note: (1) For the estimation of running time, use the following rule of thumb: 2 minutes plus $N \cdot R \cdot I \not- (T/50)$ seconds for each problem. Where N is the number of variables, R is the number of replicates, I is the number of covariates, and T is the input data in one replicate as explained under 1.d. above.

   (2) For the estimation of pages of printing, use the following rule of thumb: 3 pages plus $2^{(N-1)}$ pages for each problem. Where N is the number of variables.

b. Control Card: (One Control Card for each problem)

Col.	1,2	Problem number
Col.	3,4	Number of variables
Col.	5-7	Number of replicates
Col.	8,9	Number of covariates
Col.	10,11	01 if the transformation of variate and covariates is desired. (See the Transformation Card section below.) 00 if it is not desired.
Col.	12-16	Value of constant (c) for $\log_{10}(x + c)$ transformation in the form xx.xx. If it is not used, leave these columns blank.
Col.	17-19	Number of levels of 1st variable
Col.	20-22	Number of levels of 2nd variable
Col.	23-25	Number of levels of 3rd variable
etc.		

c. Transformation Card:

This card is prepared only when 01 is punched in Col. 10 and 11 of the Control Card. The Transformation Card has the following format:

Col.	1,2	Transformation code for variate
Col.	3,4	Transformation code for 1st covariate
Col.	5,6	Transformation code for 2nd covariate
Col.	7,8	Transformation code for 3rd covariate
etc.		

The codes referred above are:

00 no transformation

01 $\sqrt{x}$ transformation

02 $\sqrt{x} + \sqrt{x + 1}$ transformation

03 $\log_{10} x$ transformation

04 $\log_{10}(x + c)$ transformation
When this transformation is desired, the value of c must be specified in Col. 12-16 of the Control Card.

05 $\sin^{-1}\sqrt{x}$ transformation

06 $\sin^{-1}\sqrt{x/(n + 1)} + \sin^{-1}\sqrt{(x + 1)/(n + 1)}$ transformation

d. Scale Card:

Since the size of input data and the constants explained below cannot exceed 5 digits (or 6 if sign is omitted), it may be necessary to scale the data in order not to lose significant digits in data inputs. The locations of the decimal points are determined by the scale factors.

Example:	Input Data	Data Keypunched	Scale Factor
1)	312	+00312	+00000
2)	-31245000	-31245	+00003
3)	.00031245	+31245	-00008

The preparation of the Scale Card is as follows:

Col.	1-6	The scale factor as above for the variate, Y. (Sign in the first position with 5 digit scale factor)
Col.	7-12	The scale factor as above for the 1st covariate
Col.	13-18	The scale factor as above for the 2nd covariate.
etc.		

Positive sign need not be punched. Consequently, if it is not necessary to scale the data, keypunch all zeros in the proper card columns.

e. Constants Card:

This card specifies the constants used in computing the F values for the regression coefficients. If these are specified, there must be one constant for each covariate. The scale factors that are used to determine the location of the decimal points for input data also apply to these constants. Therefore, the constants are punched without decimal points and the order of magnitude of these constants must be the same as the order of magnitude of the input data.

The preparation of the Constant Card is as follows:

Col.	1-6	A sign and constant for 1st covariate
Col.	7-12	A sign and constant for 2nd covariate
Col.	13-18	A sign and constant for 3rd covariate
etc.		

If it is not necessary to specify the constants at all, leave the Constants Card blank. In this case, the program will treat them as zeros and will compute the F value.

f. Input Data Cards:

The form of input data is illustrated in the following example.

Consider the 3-variable analysis of variance design, where one variable has 2 levels, the other two variables have 3 levels each and the entire experiment is replicated twice. The data may be represented in the form:

$(y, x_1, x_2, \ldots, x_p)$ rijk

r = 1, 2 (replicates)
i = 1, 2, 3 var. 1
j = 1, 2, 3 var. 2
k = 1, 2 var. 3
p = no. of covariates ≤ 8

In preparing data for analysis every data set $(y, x_1, x_2, \ldots, x_p)$ is keypunched on a separate card in the manner described in the next paragraph. When arranging cards for processing it is necessary that the replicate index appear first as shown above, or in other words, the replicate index is to be incremented last. Other than this restriction the program labels analysis of variance variables according to the order of the data cards.

If the data are keypunched in the order

$$(y, x_1, x_2, \ldots, x_p)_{r111}$$

$$(y, x_1, x_2, \ldots, x_p)_{r112}$$

$$(\qquad\qquad\qquad)_{r121}$$

$$(\qquad\qquad\qquad)_{r122}$$

$$(\qquad\qquad\qquad)_{r131}$$

$$(\qquad\qquad\qquad)_{r132}$$

$$(\qquad\qquad\qquad)_{r211}$$

$$\vdots$$

$$(y, x_1, x_2, \ldots, x_p)_{r332}$$

then the program considers

i = 1, 2, 3 variable as variable 1
j = 1, 2, 3 variable as variable 2
k = 1, 2 variable as variable 3

Each data card is prepared using the following format

Col.	1-6	A sign and y value
Col.	7-12	A sign and x_1 value
Col.	13-18	A sign and x_2 value
	.	
	.	
	.	

The maximum number of values that can be keypunched per card is 9; one value for the y variate and eight values for the x covariates.

Leading zeros may be indicated by blanks or by punched zeros.

g. End Card:

This card will notify the program of the end of the entire job. This card has the following format:

Col. 1-10 All 9 punch

3. Order of Cards:

More than one problem can be processed consecutively. The following example illustrates the order of cards.

1) I. D. Card
2) Program Cards**
3) Control Card for problem 1
4) Scale Card
5) Constants Card
6) Input Data Cards

7) Control Card for problem 2
8) Scale Card
9) Constants Card
10) Input Data Cards
.
.
.
11) Control Card for the last problem
12) Transformation Card (if used)
13) Scale Card
14) Constants Card
15) Input Data Cards
16) End Card

** This is a prepared binary deck of 269 cards labelled BIMED # 013.

```
  CARDS FOR PROBLEM

*00451 003 005 BINED 013                                  I.D.

PROGRAM CARDS-BINARY DECK

0102003020000000002002                                    CONTROL

000000000000000000                                        SCALE

   4   12                                                 CONSTANT

   00-   12    00                                         DATA 1
   06    00    00                                         DATA 2
   13    00    00                                         DATA 3
   12    00    00                                         DATA 4
   05    00    00                                         DATA 5
   05    00    00                                         DATA 6
   14    00    00                                         DATA 7
   00    00-   12                                         DATA 8
   03    00    00                                         DATA 9
   07    00    00                                         DATA 10
   15    00    00                                         DATA 11
   10    00    00                                         DATA 12

9999999999                                                END
```

```
ANALYSIS OF COVARIANCE
PROBLEM NO.   1

NO. OF VARIABLES     2
NO. OF REPLICATES    3
NO. OF COVARIATES    1

VARIABLE     NO. OF LEVELS
    1              2
    2              2

TABLE OF VARIATION FOR VARIABLE  1
VARIATE
   120.33333     38.00000    -38.00000
COVARIATES
                 12.00000    -12.00000
                              12.00000

TABLE OF VARIATION ADJUSTED BY RESIDUALS
VARIATE
   219.66666     70.00000     30.00000
COVARIATES
                108.00000    -12.00000
                             108.00000

INVERSE OF COVARIATES MATRIX
                  0.00937      0.00104
                               0.00937

REGRESSION COEFFICIENTS
                  0.70833      0.54167


TABLE OF VARIATION FOR VARIABLE  2
VARIATE
     8.33333    -10.00000     10.00000
COVARIATES
                 12.00000    -12.00000
                              12.00000
```

```
TABLE OF VARIATION ADJUSTED BY RESIDUALS
VARIATE
   107,66666     22,00000     98,00000
COVARIATES
                108,00000    -12,00000
                             108,00000

INVERSE OF COVARIATES MATRIX
                  0,00937      0,00104
                               0,00937

REGRESSION COEFFICIENTS
                  0,30833      0,94167

TABLE OF VARIATION FOR VARIABLE  1 2
VARIATE
    75,00001     30,00000     30,00000
COVARIATES
                 12,00000     12,00000
                              12,00000

TABLE OF VARIATION ADJUSTED BY RESIDUALS
VARIATE
   174,33334     62,00000    118,00000
COVARIATES
                108,00000     12,00000
                             108,00000

INVERSE OF COVARIATES MATRIX
                  0,00937     -0,00104
                               0,00937

REGRESSION COEFFICIENTS
                  0,65833      1,04167


TABLE OF VARIATION FOR RESIDUALS
VARIATE
```

```
   99.33333    32.00000    88.00000
COVARIATES
               96.00000     0.
                           96.00000

INVERSE OF COVARIATES MATRIX
                0.01042    -0.
                            0.01042
REGRESSION COEFFICIENTS
                0.33333     0.91667

COMPUTED T VALUES
                2.82843     7.77817
RESIDUAL MEAN SQUARE        1.33333
SQRT. OF RESID, MEAN SQ.    1.15470
F ( 2,  6) =  7178.66644


TABLE OF VARIATION FOR TOTAL
VARIATE
   303.00000 . 90.00000    90.00000
COVARIATES
              132.00000   -12.00000
                          132.00000

SOURCE OF       DEGREES OF       SUMS OF       MEAN
VARIATION        FREEDOM         SQUARES       SQUARES
1                  1            134.99999     134.99999
2                  1              0.60000       0.60000
12                 1             15.00000      15.00000
WITHIN REPLICATES  6              8.00000       1.33333

THIS IS ALL THE OUTPUT FOR THIS JOB

TIME 03 35 P.M.       108 LINES OUTPUT.   TOTAL TIME OF THIS JOB   1.41  MINUTES, AS FOLLOWS
```